

09/647265

PCT/JP 00/00424

27.01.00

日本国特許庁

PATENT OFFICE
JAPANESE GOVERNMENT

EU

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日

Date of Application:

1999年 1月29日

出願番号

Application Number:

平成11年特許願第023068号

出願人

Applicant (s):

ソニー株式会社

REC'D 14 FEB 2000

WIPO

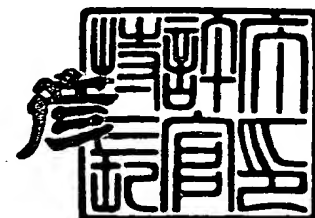
PCT

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

1999年11月19日

特許庁長官
Commissioner,
Patent Office

近藤隆彦



出証番号 出証特平11-3080175

【書類名】 特許願
【整理番号】 9900035603
【あて先】 特許庁長官殿
【国際特許分類】 G01B 11/24
【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
内

【氏名】 トビー ウォーカー

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社
内

【氏名】 松原 弘

【特許出願人】

【識別番号】 000002185

【氏名又は名称】 ソニー株式会社

【代表者】 出井 伸之

【代理人】

【識別番号】 100067736

【弁理士】

【氏名又は名称】 小池 晃

【選任した代理人】

【識別番号】 100086335

【弁理士】

【氏名又は名称】 田村 榮一

【選任した代理人】

【識別番号】 100096677

【弁理士】

【氏名又は名称】 伊賀 誠司

【手数料の表示】

【予納台帳番号】 019530

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9707387

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 データ記述方法及びデータ処理装置

【特許請求の範囲】

【請求項1】 入力データの内容の特徴を表す特徴データを記述するためのデータ記述方法であって、

記述方法を定義する単位要素となる記述子は、下位要素としての属性が定義される構造を有し、上記属性として記述子を含み得ること

を特徴とするデータ記述方法。

【請求項2】 上記入力データは、ビデオデータにおける映像データと音声データとの少なくとも一方を含むこと

を特徴とする請求項1記載のデータ記述方法。

【請求項3】 上記記述子は、上記構造の形成能力を含めて上位の記述子の機能が継承されて生成されること

を特徴とする請求項1記載のデータ記述方法。

【請求項4】 上記下位要素としての属性は、記述子と特定の定義を定義することにより構成されること

を特徴とする請求項1記載のデータ記述方法。

【請求項5】 入力データの内容の特徴を表す特徴データを生成するデータ処理装置であって、

記述方法を定義する単位要素となる記述子であり、下位要素としての属性が定義される構造を有し、上記属性として記述子を含み得る記述子から特徴データを生成する生成手段を備えること

を特徴とするデータ処理装置。

【請求項6】 上記入力データは、ビデオデータにおける映像データと音声データとの少なくとも一方を含むこと

を特徴とする請求項5記載のデータ処理装置。

【請求項7】 上記記述子は、上記構造の形成能力を含めて上位の記述子の機能が継承されて生成されたものであること

を特徴とする請求項5記載のデータ処理装置。

【請求項 8】 上記下位要素としての属性は、記述子と特定の定義を定義することにより構成されていること

を特徴とする請求項 5 記載のデータ処理装置。

【請求項 9】 上記特徴データを、上記入力データとともに、外部機器へと送信すること

を特徴とする請求項 5 記載のデータ処理装置。

【請求項 10】 入力データの内容の特徴を表す特徴データを利用するデータ処理装置であって、

記述方法を定義する単位要素となる記述子であり、下位要素としての属性が定義される構造を有し、上記属性として記述子を含み得る記述子から生成された特徴データを利用して、上記入力データの内容の特徴を復元する復元手段を備えること

を特徴とするデータ処理装置。

【請求項 11】 上記入力データは、ビデオデータにおける映像データと音声データとの少なくとも一方を含むこと

を特徴とする請求項 10 記載のデータ処理装置。

【請求項 12】 上記記述子は、上記構造の形成能力を含めて上位の記述子の機能を継承して生成されたものであること

を特徴とする請求項 10 記載のデータ処理装置。

【請求項 13】 上記下位要素としての属性は、記述子と特定の定義を定義することにより構成されていること

を特徴とする請求項 10 記載のデータ処理装置。

【請求項 14】 上記復元手段は、上記入力データの特徴を復元して上記入力データの加工されたデータを生成すること

を特徴とする請求項 10 記載のデータ処理装置。

【請求項 15】 上記特徴データを、上記入力データとともに、外部機器から受信すること

を特徴とする請求項 10 記載のデータ処理装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、入力データの内容の特徴を表す特徴データを記述するためのデータ記述方法に関する。また、本発明は、この特徴データを生成するデータ処理装置に関する。さらに、本発明は、この特徴データを利用するデータ処理装置に関する。

【0002】

【従来の技術】

例えばビデオデータに録画されたテレビ番組といった大量の異なる映像データにより構成される映像アプリケーションの中から、興味のある部分等の所望の部分を探して再生したい場合がある。

【0003】

このように、所望の映像内容を抽出するための一般的な技術としては、アプリケーションの主要場面を描いた一連の映像を並べて作成されたパネルであるストーリーボードがある。このストーリーボードは、ビデオデータをいわゆるショットに分解し、各ショットにおいて代表される映像を表示したものである。このような映像抽出技術は、そのほとんどが、例えば“G. Ahanger and T.D.C. Little, A survey of technologies for parsing and indexing digital video, J. of Visual Communication and Image Representation 7:28-4, 1996”に記載されているように、ビデオストラクチャからショットを自動的に検出して抽出するものである。

【0004】

【発明が解決しようとする課題】

ところで、上述したビデオストラクチャを抽出するには、ビデオデータの内容に応じた様々なレベルにおけるストラクチャを把握する必要があった。そのため、このようなビデオストラクチャを抽出する処理は、非常に複雑であり、ビデオストラクチャを抽出するための装置は、高度な処理能力や大容量の計算資源を有するものである必要があった。したがって、要するコストや労力を踏まえると、

家庭においてこのような処理を行うことは、非常に困難であった。

【0005】

本発明は、このような実情に鑑みてなされたものであり、ビデオデータの内容を検索してブラウジングする高度な方法を実現し、特に家庭等における種々のビデオデータの自動内容解析によるいわゆるメタ・データを生成するためのデータを記述するためのデータ記述方法を提供することを目的とするものである。また、本発明は、このようなデータを生成するデータ処理装置を提供することを目的とするものである。さらに、本発明は、このようなデータを利用して、ビデオ・コンテンツに容易でかつ個人的なアクセスを提供するメタ・データを生成するデータ処理装置を提供することを目的とするものである。

【0006】

【課題を解決するための手段】

上述した目的を達成する本発明にかかるデータ記述方法は、入力データの内容の特徴を表す特徴データを記述するためのデータ記述方法であって、記述方法を定義する単位要素となる記述子は、下位要素としての属性が定義される構造を有し、属性として記述子を含み得ることを特徴としている。

【0007】

このような本発明にかかるデータ記述方法は、構造を有する記述子の集合を用いて構造化された特徴データを記述する。

【0008】

また、上述した目的を達成する本発明にかかるデータ処理装置は、入力データの内容の特徴を表す特徴データを生成するデータ処理装置であって、記述方法を定義する単位要素となる記述子であり、下位要素としての属性が定義される構造を有し、属性として記述子を含み得る記述子から特徴データを生成する生成手段を備えることを特徴としている。

【0009】

このように構成された本発明にかかるデータ処理装置は、構造を有する記述子の集合を用いて構造化された特徴データを生成する。

【0010】

さらに、上述した目的を達成する本発明にかかるデータ処理装置は、入力データの内容の特徴を表す特徴データを利用するデータ処理装置であって、記述方法を定義する単位要素となる記述子であり、下位要素としての属性が定義される構造を有し、属性として記述子を含み得る記述子から生成された特徴データを利用して、入力データの内容の特徴を復元する復元手段を備えることを特徴としている。

【0011】

このように構成された本発明にかかるデータ処理装置は、特徴データに基づいて、入力データの内容の特徴を復元する。

【0012】

【発明の実施の形態】

以下、本発明を適用した具体的な実施の形態について図面を参照しながら詳細に説明する。

【0013】

本発明を適用した実施の形態は、例えば、映像データと音声データとからなるビデオデータといったマルチメディア・コンテンツ等を供給する放送局等における送信装置と、例えばデジタルビデオやホーム・サーバといった送信装置からのマルチメディア・コンテンツ等を受信して処理する受信装置とにより構成されるデータ送受信処理システムである。このデータ送受信処理システムは、ここでは、マルチメディア・コンテンツとしてビデオデータを扱うものとする。データ送受信処理システムの具体的な説明を行う前に、ここではまず本発明において対象とするマルチメディア・コンテンツであるビデオデータに関する説明を行う。

【0014】

本発明において対象とするビデオデータについては、図1に示すようにモデル化し、フレーム、セグメント、シーンといったレベルに階層化されたストラクチャを有するものとする。すなわち、ビデオデータは、最下位層において、一連のフレームにより構成される。また、ビデオデータは、フレームの1つ上の階層として、連続するフレームのひと続きから形成されるセグメントにより構成される

。さらに、ビデオデータは、セグメントのさらに1つ上の階層として、このセグメントを意味のある関連に基づきまとめて形成されるシーンにより構成される。さらにまた、ビデオデータは、ここでは図示しないが、関連するシーンをまとめて形成されるトピックといったさらに上の階層をも必要に応じて有するものとする。すなわち、ビデオデータは、同図におけるフレームをセグメントに、セグメントをシーンに、シーンをトピックに置き換えてモデル化することもできる。

【0015】

このビデオデータは、同期された映像及び音声のストリームの集合である。すなわち、このビデオデータにおいてフレームは、単一の静止画像である映像フレームと、一般に数十～数百ミリ秒／長といった短時間においてサンプルされた音声情報を表す音声フレームとを含むものである。

【0016】

また、セグメントは、単一のカメラにより連続的に撮影された映像フレームのひと続きからなり一般にショットと呼ばれる映像セグメントと音声セグメントとを含むものである。このセグメントは、ビデオストラクチャにおける基本となる単位である。セグメントのうち、音声セグメントは、多くの定義が可能とされ、例えば、以下に示すようなものが考えられる。例えば、音声セグメントは、一般によく知られている方法により検出されたビデオデータ中の無音期間により境界を定められて形成されるものがある。また、音声セグメントは、“D. Kimber and L. Wilcox, Acoustic Segmentation for Audio Browsers, Xerox Parc Technical Report”に記載されているように、例えば、音声、音楽、ボイド、無音等のように少数のカテゴリに分類された音声フレームのひと続きから形成されるものがある。さらに、音声セグメントは、“S. Pfeiffer, S. Fischer and E. Wolfgang, Automatic Audio Content Analysis, Proceeding of ACM Multimedia 96, Nov. 1996, pp21-30”に記載されているように、2枚の連続する音声フレーム間のある特性における大きな変化を検出する音声カット検出を用いて決定されるものがある。

【0017】

さらに、シーンは、ビデオデータの内容を意味のあるより高いレベルでまとめ

てグループ化するために、映像セグメント（ショット）検出或いは音声セグメント検出により映像フレーム或いは音声フレームを捕捉したセグメントを、例えばセグメント内の知覚的アクティビティ量といったセグメントの特性を表す特徴情報（Feature）を用いて意味のあるまとまりにグループ化したものである。

【0018】

さらにまた、トピックは、意味的に関連するシーンのシーケンスを表すものである。

【0019】

本発明を適用した実施の形態として図2に示すデータ送受信処理システムは、マルチメディア・コンテンツとしての上述したビデオデータ等の各種データを供給する送信装置10と、この送信装置10からビデオデータ等の各種データを受信する受信装置20とを備える。

【0020】

送信装置10は、同図に示すように、上述したビデオデータを保持する記憶部11と、この記憶部11からのビデオデータの内容の特徴を表してビデオストラクチャ等を記述した特徴データを生成する生成手段である特徴データ生成部12と、記憶部11からのマルチメディア・コンテンツをエンコードするエンコーダ13と、特徴データ生成部12からの特徴データをエンコードするエンコーダ14とを備える。

【0021】

また、受信装置20は、同図に示すように、送信装置10からのエンコードされたマルチメディア・コンテンツをデコードするデコーダ21と、送信装置10からのエンコードされた特徴データをデコードするデコーダ22と、特徴データに基づいてビデオストラクチャを解析し、所望の内容をユーザに提供する復元手段である検索エンジン23とを備える。

【0022】

このようなデータ送受信処理システムにおいて、送信装置10は、記憶部11に保持されたビデオデータのビデオストラクチャ等を記述する特徴データを生成し、図示しない伝送路を介して、この特徴データをビデオデータとともに受信装

置 20 へ送信する。そして、受信装置 20 は、受信した特徴データに基づき、受信したビデオデータのビデオストラクチャを回復し、例えば、上述した各シーンを代表する一連のフレームを閲覧可能としたデータであるストーリーボード等を生成する。

【0023】

このことにより、データ送受信処理システムにおいては、例えば、ユーザが受信装置 20 における検索エンジン 23 を介して、ビデオデータの中の所望の内容を見たいという要求を受信装置 20 に与えると、受信装置 20 は、検索エンジン 23 により、回復したビデオデータのビデオストラクチャを解析して、所望の内容をユーザに提供することができる。したがって、ユーザは、ビデオデータに対して容易でかつ個人的にアクセスすることが可能となり、ブラウジングを効率よく行うことができる。

【0024】

このように、ビデオデータの内容から抽出された低レベルの特徴を用いて高レベルのビデオストラクチャを抽出するための特徴データの記述内容について詳細に説明する。

【0025】

ここで、ビデオデータにおける映像と音声との内容の特徴を記述する記述方法をディスクリプションスキーム (Description Scheme) として定義する。このディスクリプションスキームは、ビデオデータにおける映像と音声との両方の成分の内容の特徴を基本的に記述するためのビデオ・コア記述法 (Video Core Description Scheme; 以下、VCS と略記する。) と、この VCS によって表される構造の視覚化を可能とする記述子であるディスクリプタにより VCS を拡張し、特にビデオデータをブラウジングするために用いるビデオ・ブラウジング記述法 (Video Browsing Description Scheme; 以下、VBS と略記する。) とからなる。

【0026】

VCS は、ビデオデータにおける様々なビデオストラクチャを捉えるために使用可能なものであって、ビデオデータ内の映像セグメント及び音声セグメント間

のストラクチャと相互関係とを把握可能とするものである。また、VCSは、ビデオデータにおける互いに類似するセグメントをまとめて形成された分類（類似性）ストラクチャと、時間的に連続したセグメントの集合である時間的セグメンテーションストラクチャと、論理的ストラクチャという3つの異なるストラクチャにおける側面を階層的に層を重ねる方法で捉え、単一の理路整然とした記述方法に統合する。さらに、VCSは、ビデオデータの低レベル及び意味論的構造の両方について、様々な観念的レベルを用いてビデオデータを記述する。さらにまた、VCSは、細部についての多層レベルにおいて、ビデオデータに関する情報にアクセスすることを可能とする。

【0027】

一方、VBSは、ビデオデータを論理的に構成して、視聴者がその内容を理解し、興味がある部分を見出すことを可能とするものである。また、VBSは、フレームのような低レベルのみではなく、シーンのような高レベルについても、類似性に基づく比較により検索を可能とするものである。

【0028】

このようなディスクリプションスキームは、その記述子であるディスクリプタの集合である。

【0029】

ディスクリプタは、その属性を定義することによって、構造を有することが可能となっており、下位要素として自らを含めた他のディスクリプタと特定の値とを定義することにより構成される。ディスクリプタは、例えば図3に示すような記述フォーマットを有している。すなわち、ディスクリプタの記述フォーマットは、そのディスクリプタ名とディスクリプタコードとを示すディスクリプタ (Descriptor) と、このディスクリプタの上位クラスのディスクリプタを示すスーパークラス (Superclass) と、下位要素のディスクリプタ或いは数値の種類を示すフィーチャ (Feature) と、下位要素のディスクリプタ或いは数値のタイプを示すフィーチャタイプ (Feature Type) と、下位要素のディスクリプタ或いは数値の例を示す実例値 (Example Value) とからなる。

【0030】

ここで、上述したスーパークラスとして示される上位クラスのディスクリプタとは、いわゆるオブジェクト指向表現において、あるディスクリプタが継承される元のディスクリプタを指す。すなわち、ディスクリプタは、ここでは、オブジェクト指向表現により記述され、元のディスクリプタの機能を継承している。このことにより、ディスクリプタは、より簡便に高度な機能を有するものへと拡張することができ、このディスクリプタを用いて記述される特徴データは、情報量の少ないより簡便な構成により作成され得る。このディスクリプタの継承関係は、ここでは、図4に示すようなものであるとする。

【0031】

ディスクリプタは、同図に示すように、これらのVCS及びVBSのいずれかに属し、1つの体系を形成する。

【0032】

すなわち、VCSに属するディスクリプタとしては、オーディオ・ビジュアル・エレメント (Audio Visual Element; 以下、AVエレメント又はAVElementと略記する。) と、オーディオ・ビジュアル・セグメント (Audio Visual Segment; 以下、AVセグメント又はAVSegmentと略記する。) と、オーディオ・ビジュアル・アブストラクト (Audio Visual Abstract; 以下、AVアブストラクト又はAVAbstractと略記する。) と、オーディオ・ビジュアル・プロトタイプ (Audio Visual Prototype; 以下、AVプロトタイプ又はAVPrototypeと略記する。) と、オーディオ・ビジュアル・ハイアラキ (Audio Visual Hierarchy; 以下、AVハイアラキ又はAVHierarchyと略記する。) と、クラスタ (Cluster) と、コンポジッド・セグメント (Composite Segment) と、オーディオ・ビジュアル・グラフ (Audio Visual Graph; 以下、AVグラフ又はAVGraphと略記する。) と、オーディオ・ビジュアル・グラフ縁 (Audio Visual Graph Edge; 以下、AVグラフ縁又はAVGraphEdgeと略記する。) と、シグネチャ (Signature) と、テンポラル・シグネチャ (Temporal Signature) と、アクティビティ・レベル (Activity Level) と、ウエイト (Weight) と、ウエイト・ベクトル (Weight Vector) と

、カラー (Color) と、カラー・スペース (Color Space) と、コンセプト (Concept) と、カラー・シグネチャ (Color Signature) とがあり、参照情報として XML, RDFがある。

【0033】

一方、VBSに属するディスクリプタとしては、シノプシス (Synopsis) と、オーディオ・フレーム (Audio Frame; 以下、音声フレームと略記する。) と、ビジュアル・フレーム (Visual Frame; 以下、映像フレームと略記する。) と、ブラウジング・ノード (Browsing Node) と、ショット (Shot) と、シーン (Scene) と、トピック (Topic) と、プログラム (Program) と、ハイライト (Highlight) とがある。

【0034】

ここで、同図中において、△が付随しているディスクリプタは、ディスクリプタ間の継承関係における上位クラスのディスクリプタであることを示し、◇が付随しているディスクリプタは、それと接続しているディスクリプタが属性を表すディスクリプタであることを示しているものとする。すなわち、ディスクリプタの継承関係において、AVセグメントと、AVアブストラクトと、AVハイアラキと、クラスタと、AVグラフとに対する継承元のディスクリプタは、AVエレメントである。また、ディスクリプタの継承関係において、AVグラフの属性を表すディスクリプタは、AVグラフ縁である。同様に、他のディスクリプタについても、継承元のディスクリプタと、属性を表すディスクリプタとを示すことができる。

【0035】

以下、これらのディスクリプタについて順次説明していく。なお、以下の議論の中では、ビデオデータは、1つの映像ストリームと1つの音声ストリームとが存在するものとみなして説明を進めるが、ディスクリプションスキームは、ビデオデータ中に任意の数の映像ストリームと音声ストリームとが含まれているようなさらに一般的な場合も取り扱うことができるものである。

【0036】

まず、VCSに属するディスクリプタとして、AVエレメント・ディスクリプ

タについて説明する。映像音声ストリームの内容解析を行うための論理ユニットを表す抽象的なディスクリプタとしてAVエレメントを導入する。このAVエレメント・ディスクリプタの定義は、映像フレームのような基本的な解析ユニット、ショットやシーンのような高レベルの構造ユニット、例えばある俳優が登場する全てのフレームといったある分類法に基づく任意のグループ化を共に包含するように意図的に未確定とされている。

【0037】

このようなAVエレメント・ディスクリプタの記述フォーマットは、図5に示すようなものになる。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における(AVE)は、ディスクリプタコードを示している。このAVエレメント・ディスクリプタは、映像音声ストリームの内容を記述するための抽象的なエレメントを定義する。ディスクリプションスキームにおいて、映像音声ストリームの内容のある側面を記述する他の全てのノードは、このAVエレメント・ディスクリプタを引き継いでいる。AVエレメント・ディスクリプタは、ビデオデータの実際の部分を指してもよく、或いはビデオデータの内容を記述するために作られた抽象的或いは解析的構成物を指してもよい。

【0038】

ディスクリプションスキームにおいて、ビデオデータ中のノードは、1つのノードが他の記述エレメントを参照する必要がある場合には、常に、このタイプを用いる。このことにより、異なる内容解析ツールは、自由に相互混合される。また、この方法により、抽象化が低レベルから高レベルへと上昇していく。最も低いレベルにおいては、AVエレメントは、生のメディアに近いエレメントを表し、より高いレベルにおいては、コンテンツ・レベルの特徴を記述する。両方の場合において、記述用のツールの同じセットが、階層的なやり方で使用でき、かつ相互リンクされ得る。

【0039】

同様にして、特徴値 (Feature) が、それらの起源である値のみならず、解析的に得られる構成物にも自由に付加できる。例えば、カラー・ヒストグラムが単

一の画像だけではなく、ショットのグループを表すAVエレメントに付加される。

【0040】

つぎに、AVセグメント・ディスクリプタについて説明する。ディスクリプションスキームにおいて最も重要なAVエレメントは、AVセグメント・ディスクリプタである。AVセグメントは、少なくとも1つ以上の映像音声ストリームを含むビデオデータのストリームの区切られた時間間隔を示す。ストリームの集合は、映像、音声又は両方のタイプのストリームを含んでいてもよい。この時間間隔は、ビデオデータに付随された統一的な時間により示してもよく、または、ビデオデータに統一的に付けられたフレーム番号等により指定してもよい。AVセグメントの一例としては、1つのビデオフレーム或いは完全なカメラのショットを表す一連のフレームの内容がある。

【0041】

このようなAVセグメント・ディスクリプタの記述フォーマットは、図6に示すようなものになり、AVセグメント・ディスクリプタは、スーパークラスとしてAVエレメント・ディスクリプタを有する。また、AVセグメント・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがタイムコード (Timecode) であり、ビデオデータ内でのこのセグメントのスタートポイントを示すスタートポイント (Start Point) と、そのタイプがタイムコード (Timecode) であり、ビデオデータ内でのこのセグメントのエンドポイントを示すエンド・ポイント (End Point) と、そのタイプが指示されているストリームの集合 (set of streamref) であり、ビデオデータ内でそのセグメントによって指示されているストリームの集合を示すストリームズ (Streams) と、そのタイプがAVアブストラクトの集合 (set of AVAbstract) であり、このセグメントの映像音声上の内容の要約を示す0個以上の特徴を示し、実例値としてキーフレーム (Key frame) が挙げられるアブストラクト (Abstract) と、そのタイプがコンセプトの集合 (set of Concept) であり、このセグメントの内容の映像音声上の意味内容を特徴付ける0個以上の注釈を示し、実例値としてこのセグメントの内容を記述するRDF文書が挙げられるコンセプト (Concept)

t) とを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (A V S) は、ディスクリプタコードを示している。

【0042】

A V セグメントは、それ自体としては、それ自体の中に或いはそれ自体がその抽象的なレベル、範囲或いは意味を有していない。A V セグメントは、例えば、単一の映像フレームのような低レベルユニットや、テレビ番組中におけるコマーシャルのような高レベルユニットの両方を表すように用いることができる。

【0043】

映像ストリーム又は音声ストリームにおいて可能な最小の A V セグメントは、映像音声フレームと呼ばれる。すなわち、上述したように、映像ストリームについては、映像フレームは、単一の静止画像であり、音声ストリームについては、音声フレームは、通常、解析を可能とするために、単一のサンプルではなく、音声サンプルの短いシーケンスである。ディスクリプションスキームにおいては、フレームよりもセグメントに関心があるため、映像音声フレームは、音声セグメントの最も単純なノードとして取り扱われる。A V セグメントのその他の例としては、例えばショットやシーン等がある。

【0044】

なお、ディスクリプションスキームにおいては、プレースホルダーとしてタイムコード・ディスクリプタを用いる。唯一の必要条件は、それが継続時間の計測を可能とし、かつビデオデータ内の全てのストリームについて共通のタイムスケールを有する数量スケールであることである。具体例としては、ビデオデータの開始から時間を秒単位で計測することである。同様に、streamref ディスクリプタは、マルチ・ストリーム・マルチ・メディア文書内の単一ストリームへのリファレンスを表すためのディスクリプタ用のプレースホルダーである。

【0045】

つぎに、A V アブストラクト・ディスクリプタについて説明する。この A V アブストラクト・ディスクリプタの記述フォーマットは、図 7 に示すようなものになり、A V アブストラクト・ディスクリプタは、スーパークラスとして A V エレ

メント・ディスクリプタを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (ABST) は、ディスクリプタコードを示している。

【0046】

AVアブストラクト・ディスクリプタは、AVエレメントの内容の要約を示す。例えば、代表となるようなこの下位のエレメント（例えばこのエレメントがシーンを指すならば、下位のエレメントはショットとなる。）を選択し、これを要約とすることができる場合もある。理想的には、この要約に用いるAVエレメントは、2つのプロパティ (Property) を含む。すなわち、第1のプロパティとしては、このAVエレメントの重要又は適切な内容を代表するものである。第2のプロパティとしては、このAVエレメントの特徴を示すものであり、他のAVエレメントとこのAVエレメントとの違いを示すものである。

【0047】

この要約は、ビデオデータの実際の内容の代替として使用できるものである。ディスクリプションスキームにおいて、このAVアブストラクト・ディスクリプタの最も重要な使用方法は、この要約を比較するだけで類似のAVエレメントを検索することである。すなわち、この要約とは、AVエレメントの検出、比較、検索等を行うためのキーとしての役割を果たすものである。

【0048】

つぎに、AVプロトタイプ・ディスクリプタについて説明する。このAVプロトタイプ・ディスクリプタの記述フォーマットは、図8に示すようなものになり、AVプロトタイプ・ディスクリプタは、スーパークラスとしてAVアブストラクト・ディスクリプタを有する。また、AVプロトタイプ・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがストリング (String) であり、このAVプロトタイプ・ディスクリプタの特性を示すとともに、このAVプロトタイプ・ディスクリプタが示すAVエレメントのタイプを示し、実例値としてショット (Shot) やシーン (Scene) や映像フレーム (VisualFrame) が挙げられる種別 (Kind) と、そのタイプが重みベクトル (Weight Vector; ウェイト・ベクトル) であり、プロトタイプの各特徴につ

いて、関連性或いは重要性の重みを表す n 次元ベクトルを示す重み (Weights; ウェイト) と、その他必要に応じて、そのタイプがディスクリプタ 1, ..., n (Descriptor1, ..., Descriptorn) であり、第 1 乃至第 n 特徴の実例ディスクリプタ値を示す特徴 1, ..., n (Feature1, ..., Feature n) とを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (AVP) は、ディスクリプタコードを示している

【0049】

AVプロトタイプは、AV要素の概略的な表現である。特に、ディスクリプションスキームにおいては、このAVプロトタイプによって、あるタイプのAV要素のプロトタイプ的な例 (instance) をディスクリプタ値の集合として記述することができ、その後具体的に各特徴の関連性或いは重要性を重み付けで定義する。別の見方をすれば、あるプロトタイプ・要素は、非常に単純な実例による照合を行うためのものであり、指定された種類のAV要素を、その相当する特徴値 (Feature Value) で記述する。

【0050】

プロトタイプは、通常は解析による構築物であり、AV要素のあるグループに対して、具体的な代表的AV要素を対置させることが必要な場合に使用される。例えば、あるAV要素が、あるAV要素のグループにどの程度類似しているかを決定したい場合には、その要素のプロトタイプと、グループのプロトタイプの類似性を比較することができる。

【0051】

ここで、図9に示すように、例えばある単一のショット中の全ての画像を考え、この画像グループについて、プロトタイプとしての画像を定義したいと考える。この場合には、実際の画像を用いる代わりに、まず画像についての異なる特徴値、例えば色彩とテクスチャの範囲を求める。次に、それぞれの特徴について、例えば平均値のような典型的な値を決定する。この値が、プロトタイプ中の特徴値となる。そして、ある特徴がショットにおけるその集合について、どの程度関連性があるかを決定する。例えば、ショットにおいて、色彩は非常に重要である

が、テクスチャはあまり関連性がない場合がある。このような各特徴に関する関連性或いは重要性の相違は、各特徴に割り当てられる重み付けによって表され、値の大きさが特徴値についてのこのプロトタイプに対する関連性を示している。

【0052】

このように、AVプロトタイプ・ディスクリプタは、例えばショットやシーンといったビデオデータのストラクチャ解析のために用いるセグメントを代表するフレームを記述する。例えば、後述するように、AVプロトタイプにより記述されたショットを代表する映像フレームを用い、複数のショットにおける特徴値として、それぞれの代表フレームに関する特徴値を抽出して比較し、時間的に近接して類似したショットを検出して繰り返しまとめ、最終的にシーンを生成することができる。なお、ディスクリプションスキームにおいて、ユーザがビデオデータをブラウジングするために表示する代表フレームは、AVプロトタイプ・ディスクリプタではなく、後述するクラスタ・ディスクリプタにより記述される。

【0053】

つぎに、AVハイアラキ・ディスクリプタについて説明する。ビデオデータにおいて互いに類似した映像セグメントや音声セグメントを一緒にグループ化することは、内容の検索やブラウジングのためにビデオデータを組織化することに関して基本的な必要条件である。そこで、ディスクリプションスキームにおいては、ツリーに基づく階層を用いることにより、細部の様々なレベルにおいて、ある映像セグメントや音声セグメントの集合のグループ化を表すことを提案する。この階層的分類は、以下の特徴を提供する。すなわち、第1の特徴としては、細部の様々なレベルでアクセスを可能とすることであり、第2の特徴としては、内容を要約することである。また、第3の特徴としては、階層的分類法を表すことであり、第4の特徴としては、高レベルの層は、低レベルでの知覚による類似性ではなく、近い意味論的分類を捉えることができることである。

【0054】

階層の各ノードは、AVエレメントのある集合を表す。すなわち、階層の各ノードは、直接このノードに属するものと、少なくとも1つ以上のそのノードの子ノードに繰り返し含まれるものがある。この階層における各分類によりこのハ

イアララキ (Hierarchy) の複数のメンバー又は代表するメンバーを要約された特徴は、例えば、ディスクリプタが n 次元ベクトルによって表される場合には、ディスクリプタ値のセントロイド (centroid) が良好な要約となる。

【0055】

AVハイアララキ・ディスクリプタの記述フォーマットは、図10に示すようなものになり、AVハイアララキ・ディスクリプタは、スーパークラスとしてAVエレメント・ディスクリプタを有する。また、AVハイアララキ・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがクラスタ (Cluster) であり、階層のルート (起点) を示すルート (Root) と、そのタイプがストリング (String) であり、階層の種類を示す種別 (Kind) とを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (HIER) は、ディスクリプタコードを示している。

【0056】

このようなAVハイアララキ・ディスクリプタは、関連するAVエレメント又は類似するAVエレメントを1つの階層構造にしてグループとしてまとめた結果そのものを示すものである。

【0057】

つぎに、クラスタ・ディスクリプタについて説明する。このクラスタ・ディスクリプタの記述フォーマットは、図11に示すようなものになり、クラスタ・ディスクリプタは、スーパークラスとしてAVエレメント・ディスクリプタを有する。また、クラスタ・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがAVエレメントの集合 (set of AVElement) であり、このグループ内に直接含まれるAVエレメントの集合を示すメンバー (Members) と、そのタイプがAVアブストラクトであり、このクラスタに含まれるセグメントの集合から抽出されたこのAVエレメントの集合の代表的なメンバーを表す値を示す構造 (Struct) とを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (CN) は、ディスクリプタコードを示している。また、全てのメンバーの集合

は、回帰的に見出すことができる。

【0058】

このようなクラスタ・ディスクリプタは、関連するAVエレメント或いは類似するAVエレメントと一緒にグループ化して、1つのクラスを形成することを表すものである。この定義は、回帰的であって、あるクラスタは、メンバーとして他のクラスタを含む可能性がある。その場合には、その結果は、階層的な分類構造となる。

【0059】

このクラスタ・ディスクリプタを用いることによって、例えば図12に示すように、セグメント間の類似性 (similarity) を基にしたセグメントのクラスタリングを行うことができ、このクラスタリングの結果を記述することができる。具体的には、クラスタ・ディスクリプタを用いることによって、後述するように、類似性を基にしてショットのクラスタリングを行う。

【0060】

クラスタ・ディスクリプタの定義においては、階層的な分類の構造と分類ノードに含まれるメンバーとに対して、殆ど制約を課していない。そこで、多くの場合において、次のような制約が分類ツリーに追加される。

【0061】

すなわち、第1の制約としては、子ノードの要素数の制約がある。すなわち、各クラスタ・ディスクリプタで許容される子ノードの数は、制約されてもよい。例えば、要素数を2に限定する場合には、2進ツリー構造を構成することができる。また、各ノードが固定要素数を有する場合には、2進ツリー、4進ツリー、或いは8進ツリーに基づく構成を規定することができる。

【0062】

また、第2の制約としては、分割の制約がある。すなわち、ある映像セグメント及び／又は音声セグメントを分割してクラスタに分ける場合には、そのセグメントの要素は、重複して2つのクラスタの要素となることはない。また、その要素全ては、必ず1つのクラスタの要素になっている。

【0063】

さらに、第3の制約としては、レベルの制約がある。上述した定義により、いかなる高さのツリーでも構成することが可能となる。例えば、ツリーが1レベル以上でなければならないという必要性は存在しない。ただし、このような制限がある分類法も存在する可能性がある。

【0064】

このような類似性ツリーを構成する方法としては、多数のものが知られており、例えば、凝集階層クラスタリング法 (agglomerative hierarchical clustering methods) 等がある。

【0065】

また、階層的分類をやや異なる角度から見ることによって、ディスクリプタ値を求めるための階層的量子化法を表わすものとして階層的分類を解釈することができる。特に、最も類似した分類グループに任意のディスクリプタ値を割り当てることによって、そのディスクリプタに量子化値を選択している。そして、グループに関連するAVプロトタイプ・ディスクリプタ値は、量子化値として利用される。

【0066】

例えば、 n 次元ベクトル空間によって表されるディスクリプタのためのベクトル数量化法を考える。ある値の集合を数量化するために用いられるベクトルは、量子化のためのあるコード・ブックを生成するために、例えば一般化されたLloydアルゴリズム (Lloyd algorithm) やKohonen自己組織化マップ (Kohonen Self-Organizing Maps) といった何らかのクラスタリング・アルゴリズムを用いて、特徴空間を領域の集合に分割することによって通常見出される。1つのクラスターを各コード・ブックのエントリーに割り当てて、そのコード・ブック値をプロトタイプ値として用いる場合には、階層的分類を用いて、VQコード・ブックを表すことができる。

【0067】

つぎに、コンポジッド・セグメント・ディスクリプタについて説明する。階層的分類は、AVセグメントのグループ間の類似性を、セグメント間の時間的順序

に関心を抱くことなく捉える。そこで、ビデオデータの時間的層組織を捉えるコンポジッド・セグメントの観念を導入する。コンポジッドは、例えば図13に示すように、AVエレメントの一連の要素から形成されるセグメントを代表するものである。

【0068】

コンポジッド・セグメント・ディスクリプタの記述フォーマットは、図14に示すようなものになり、コンポジッド・セグメント・ディスクリプタは、スーパークラスとしてAVセグメント・ディスクリプタを有する。また、コンポジッド・セグメント・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがAVセグメントの集合 (set of AVSegment) であるコンポーネント (Components) を有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (SN) は、ディスクリプタコードを示している。また、このコンポジッド・セグメント・ディスクリプタは、特殊な種類のAVセグメント・ディスクリプタであるため、コンポーネントの定義は、セグメンテーションが繰り返し生起することを可能とする。

【0069】

コンポジッド・セグメント・ディスクリプタは、AVセグメントのシーケンスを、互いに類似するサブシーケンス或いは関連するサブシーケンスに分割することを表す。このコンポジッド・セグメント・ディスクリプタを用いることによって、ビデオデータのストラクチャをツリー構造として記述することができる。具体的には、コンポジッド・セグメント・ディスクリプタを用いることによって、後述するように、類似するセグメントをグループ化してクラスタリングした結果を、ツリー構造として記述したり、シーンをさらにクラスタリングして、ビデオデータのストラクチャをツリー構造として記述する。

【0070】

つぎに、AVグラフ・ディスクリプタ及びAVグラフ縁・ディスクリプタについて説明する。ビデオデータの論理的ストラクチャを捉えるために、AVグラフの概念を導入する。メディア構造グラフは、有向グラフであり、そこでは各頂点

が、音声セグメントのようなAVエレメントを表し、また各縁が、あるエレメントから他のエレメントへの遷移を表している。そして、2つの特別な頂点が、スタート頂点及びエンド頂点として指定される。

【0071】

AVグラフ・ディスクリプタの根底にある観念は、形式言語を記述するための何らかの単純なツールを、ビデオデータの論理的ストラクチャを記述するために応用することが可能であるということである。例えば、形式言語がアルファベットからとられた記号のシーケンスの観念に基づくのと同様に、ビデオデータのストリームも、AVエレメントのシーケンスとして見ることができる。AVグラフは、ステートをAVエレメントで置き換えた有限ステート・オートマトンの1変種と考えることができる。図15(A)に示すような会話場面を形成する一連のショットのための、2つの異なるメディア構造グラフを図15(B)及び図15(C)に示す。同図(B)に示す例は、単純な直線的構造を示すものであり、同図(C)に示す例は、話し手に応じて交代するショットという特徴を有する高レベルの交代構造を示すものである。

【0072】

このようなAVグラフ・ディスクリプタの記述フォーマットは、図16に示すようなものになり、AVグラフ・ディスクリプタは、スーパークラスとしてAVエレメント・ディスクリプタを有する。また、AVグラフ・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがAVグラフ縁の集合(set of AVGraphEdge)であり、このグラフ中の縁の集合を示す縁(Edges)と、そのタイプがストリング(String)であり、グラフの種類を示す種別(Kind)とを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における(AVG)は、ディスクリプタコードを示している。

【0073】

このAVグラフ・ディスクリプタは、AVエレメントの対の間の異なる種類の関係を示すラベルの付いた縁(Edge)を有するグラフに、それらとともに結合することを記述する。AVグラフは、AVエレメントを指すノードと、各ノードを

リンクするラベルを付けられた縁とを有するラベルを付けられたグラフである。

【0074】

このAVグラフ・ディスクリプタによって、ビデオデータの繰り返し構造が記述できるようになる。具体的には、AVグラフ・ディスクリプタによって、後述するように、検出したシーンの内容を、そのシーンの繰り返し構造として記述し、ビデオデータ全体のストラクチャ解析を行う場合に利用する。

【0075】

また、AVグラフ縁・ディスクリプタの記述フォーマットは、図17に示すようなものになり、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがAVエレメント (AVElement) であり、この縁のソースエレメントであることを示すソース (Source) と、そのタイプがAVエレメント (AVElement) であり、この縁の目標エレメントであることを示すターゲット (Target) と、そのタイプがストリング (String) であり、2つのエレメントにおける縁によって結合された2つのエレメントの間の関係の種類を表すラベル (Label) とを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (AVEDGE) は、ディスクリプタコードを示している。

【0076】

このようなAVグラフ縁・ディスクリプタは、2つのAVエレメントの間の有向結合 (directed connection) を表す。

【0077】

全てのAVグラフ縁は、特定のグラフに属する。また、同じAVエレメントが、異なるグラフに属することがあり得る。さらに、2つの同じAVエレメントが、異なるAVグラフ内でともにリンクされることもあり得る。

【0078】

AVグラフの定義は、ノードが参照できるAVエレメントのタイプについて、何らの特定の制約をも課さない。特にノードは、AVプロトタイプにリンクされることができる。そのような場合には、そのグラフは、AVプロトタイプ・グラフと称され、ビデオデータのストリームのストラクチャのプロトタイプとしての

例 (prototypical instance) を表す。例えば、このような構造を用いて、テレビのニュース番組におけるストラクチャを表すことができる。

【0079】

また、AVグラフは、1つのグラフ中のノードが、実際には他のグラフであるような入れ子になったグラフの階層を形成することができる。これにより、AVグラフは、マルチプル・レベルの表現の構造を捉えることが可能となる。

【0080】

つぎに、シグネチャ・ディスクリプタについて説明する。このシグネチャ・ディスクリプタの記述フォーマットは、図18に示すようなものになり、シグネチャ・ディスクリプタは、スーパークラスとしてAVアブストラクト・ディスクリプタを有する。また、シグネチャ・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがウエイト・ベクトル (WeightVector [n]) であり、重みのn次元ベクトルであることを示すウエイト (Weights) と、そのタイプがAVエレメント (AVElement [n]) であり、それぞれの重み付けに対するAVエレメントであることを示すとともに、重みとAVエレメントとが1対1に対応することを示すメンバー (Members) と、そのタイプがSTRING (String) であり、シグネチャの種類を示すとともに、同じ種類のシグネチャのみが互いに比較できることを示す種別 (Kind) とを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (SIG) は、ディスクリプタコードを示している。

【0081】

このシグネチャ・ディスクリプタは、代表的な複数のAVエレメントと、それぞれのエレメントの重み付けの値 (weights) の組み合わせにより、AVエレメントの静的な内容を代表するものである。

【0082】

1つのシグネチャは、一般化したヒストグラムとして表わすことが可能である。シグネチャが対象とするAVエレメントは、カテゴリ要素の分布状況、すなわちシグネチャのメンバーのAVエレメント・ディスクリプタの分布として表現できる。重み付けは、特定のクラスに割り当てられた分布の部分或いは確率分類ベ

クトルを示す。その場合に、各重みは、ディスクリプタ値が対応するクラスのメンバーである確率を表す。

【0083】

シグネチャの第1の例としては、ショット・キーフレーム (Shot Key Frames) がある。ショット・キーフレームは、コンテンツから抽出されたショットを代表するするフレームの組である。各キーフレームの重みは、そのキーフレームに一致した (最も類似している) フレームの割合を示す。

【0084】

また、シグネチャの第2の例としては、プログラム・シグネチャ (Program Signature) がある。プログラム・シグネチャは、あるプログラムの特徴を強く表わすショットの組である。例えば、あるニュースプログラムを考えたとき、オープニングのロゴのあるショット、キャスターのショット、最後のロゴのショット等がそのニュースプログラムの特徴を一番表わしたものとして使用できる。

【0085】

このように、シグネチャ・ディスクリプタは、ビデオデータの特徴抽出時における方式やパラメータを指定する。具体的には、シグネチャ・ディスクリプタを用いることによって、後述するように、コンテンツの種類に対して最適な特徴抽出方法を選択した場合にその方法を記述する。

【0086】

つぎに、テンポラル・シグネチャ・ディスクリプタについて説明する。このテンポラル・シグネチャ・ディスクリプタの記述フォーマットは、図19に示すようなものになり、テンポラル・シグネチャ・ディスクリプタは、スーパークラスとしてAVアブストラクト・ディスクリプタを有する。また、テンポラル・シグネチャ・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがウエイト・ベクトル (WeightVector) であり、ノードの各セクションが占める相対的継続時間を示すタイム (Time) と、そのタイプがAVエレメントのシーケンス (seq of AVElement) であり、ノードの特徴を示す映像及び音声のシーケンスであることを示すメンバー (Members) と、そのタイプがストリング (String) であり、シグネチャの種類を示すとともに、同

じ種類のシグネチャのみが互いに比較できることを示す種別 (Kind) とを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (TS) は、ディスクリプタコードを示している。

【0087】

このようなテンポラル・シグネチャ・ディスクリプタは、テンポラル・シグネチャ・ディスクリプタ・メンバーである複数の代表的なAVエレメントのシーケンスにより、当該テンポラル・シグネチャ・ディスクリプタが対象とするAVエレメントの動的な内容を代表するものである。映像及び音声のシーケンスの各メンバーは、対象とするAVエレメントの1つの状態を表わしている。それぞれの映像及び音声のシーケンスに対しては、重み付けが与えられており、このシーケンスに対応する状態の相対的な長さを示している。

【0088】

対象とするAVエレメントに必要な映像及び音声のシーケンスの数は、AVエレメントそれ自体の変化の程度と内容の細かさとの両方に依存するはずである。一番少ない例においては、変化が少ない場合はこれを1つの映像及び音声のシーケンスとして扱うことができる。また、一番多い例においては、対象とするAVエレメントの全てのサブエレメントから映像及び音声のシーケンスになる場合もある。これは、ショットの中の全てのフレームが映像及び音声のシーケンスとなる場合である。

【0089】

テンポラル・シグネチャの第1の例としては、ショット・テンポラル・シグネチャ (Shot Temporal Signature) がある。このショット・テンポラル・シグネチャは、ショットの動きの中の重要な時間を示すキーフレームのシーケンスであって、ショットの特徴を表わす。例えば、ショットからプレイヤーのズーム場面がこれに当たる。

【0090】

また、テンポラル・シグネチャの第2の例としては、テレビジョン・プログラム・テンポラル・シグネチャ (Television Program Temporal Signature) がある。このテレビジョン・プログラム・テンポラル・シグネチャは、プログラムか

ら抜き出したそのプログラムの特徴を表わすキーショットの組である。例えば、ニュースプログラムの場合には、代表的なロゴのあるショット、キャスターのショット、スポーツのショット、天気予報等が、そのニュースプログラムの特徴を一番表わしたものとして使用できる。

【0091】

上述したシグネチャとこのテンポラル・シグネチャとの両者は、AVエレメントを重み付けされた映像及び音声のシーケンスにより記述する。しかしながら、シグネチャとは異なり、テンポラル・シグネチャは、時間的順序を有している。テンポラル・シグネチャは、AVエレメントが時間の経過にともなって通過する状態のシーケンスを記述する。重み付け因子は、各クラス内でどの程度の（相対的）時間がかかったかを示す指標である。すなわち、クラスの順序は、時間的順序に対応する。

【0092】

このようなテンポラル・シグネチャ・ディスクリプタによって、ビデオデータの特徴抽出時に仮の特徴抽出方式やパラメータを指定する。具体的には、テンポラル・シグネチャ・ディスクリプタは、後述するように、コンテンツの種類に対して最適な特徴抽出方法を選択する場合において、最適解を試行錯誤により探し出す過程にいて使用する。

【0093】

つぎに、オーディオ・ビジュアル・ディスクリプタとして、アクティビティ・レベル・ディスクリプタと、ウェイト・ディスクリプタと、ウェイト・ベクトル・ディスクリプタと、カラー・ディスクリプタと、カラー・スペース・ディスクリプタと、コンセプト・ディスクリプタと、カラー・シグネチャ・ディスクリプタとについて説明する。これらのディスクリプタに関する記述は、全てのAVエレメントに適用されるものであり、記述フォーマット中の適用対象 (Applies to) の欄にその旨を記載している。

【0094】

アクティビティ・レベル・ディスクリプタの記述フォーマットは、図20に示すようなものになり、アクティビティ・レベル・ディスクリプタは、下位要素の

ディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがリアル (Real) であり、アクティビティを示す 0 以上 1 以下の実数であって、0 はアクティビティがないことを示すとともに、1 はアクティビティが最大であることを示すアクティビティ・レベル (ActivityLevel) を有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (ACT) は、ディスクリプタコードを示している。

【0095】

このアクティビティ・レベル・ディスクリプタは、AV要素内のオーディオ・ビジュアル・アクティビティ・レベルの程度を表す。アクティビティ・レベル・ディスクリプタを生成するために使用される方法は、アプリケーションの必要に依存する。

【0096】

例えば、映像シーケンスを考える。映像シーケンス中のアクティビティの量は、様々な方法で測定することができ、シーケンス内の隣接フレーム間の平均距離やシーケンス内で生起する全体的な運動の量等により測定することができる。

【0097】

このアクティビティ・レベル・ディスクリプタによって、後述するように、映像及び音声の特徴が記述され、ビデオデータの内容の変化の大きさを測るパラメータとして利用できる。

【0098】

また、ウエイト・ディスクリプタの記述フォーマットは、図 21 に示すようなものになり、ウエイト・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがリアル (Real) であり、この AV要素に付加された重みの量を示す 0 と 1 との間の実数であるウエイティング (Weighting) を有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (WGT) は、ディスクリプタコードを示している。

【0099】

このウエイト・ディスクリプタは、AV要素の相対的重み付けを表す。

重み付けの意味は、それが使用される文脈によってのみ決定される。

【0100】

さらに、ウェイト・ベクトル・ディスクリプタの記述フォーマットは、図22に示すようなものになり、ウェイト・ベクトル・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがウェイト (Weight[n]) であり、Weightのn次元ベクトルであるウェイト (Weights) と、そのタイプがブール (Boolean) であり、ベクトル内のWeightが規格化された値か否かを示すフラグである規格化 (Normalized) とを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (WV) は、ディスクリプタコードを示している。また、Weightへは、制約条件を課することができる。

【0101】

このウェイト・ベクトル・ディスクリプタは、AV要素に適用される重み付けのn次元ベクトルを表す。このウェイト・ベクトル・ディスクリプタによって、ビデオデータの特徴抽出する際の又は特徴抽出した結果のパラメータの重み付けした値を指定又は記述する。具体的には、ウェイト・ベクトル・ディスクリプタによって、後述するように、複数の類似性を示すパラメータ同士に対し、類似性を判定する際の各パラメータに対する判定貢献度を指定する。

【0102】

さらにまた、カラー・ディスクリプタの記述フォーマットは、図23に示すようなものになり、カラー・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがリアル (Real[3]) であり、

[0.0, 1.0] の範囲で表現された色彩要素値を示すカラー (Color) を有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (COL) は、ディスクリプタコードを示している。また、色彩は、実数を要素とする3次元ベクトルとしてモデル化されており、各要素は、0から1までの値をとる。

【0103】

このカラー・ディスクリプタは、色彩を3つの異なる実数で表現する。各値は

、色彩空間における 3 軸に相当する。

【0104】

また、カラー・スペース・ディスクリプタの記述フォーマットは、図 24 に示すようなものになり、カラー・スペース・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがストリング (String) であり、色彩空間の名称を示すカラー・スペース (ColorSpace) を有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (CS) は、ディスクリプタコードを示している。

【0105】

色彩は、様々な異なる色彩空間で符号化することができる。このカラー・スペース・ディスクリプタは、カラー・ディスクリプタの意味をどう解釈するかを決定する。

【0106】

さらに、コンセプト・ディスクリプタの記述フォーマットは、図 25 に示すようなものになり、コンセプト・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプが RDF ノードや XML ノード等であり、例えば RDF ノードや XML ノードに基づいた記述へのリンクと、いったこの AV エLEMENT の概念へのリンクを示すコンセプト (Concept) を有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (CPT) は、ディスクリプタコードを示している。

【0107】

このコンセプト・ディスクリプタは、AV エLEMENT を、その観念的内容 (conceptual contents) を記述する表現にリンクさせる。ディスクリプションスキームにおいては、AV エLEMENT を観念的に記述するために、どのような形式論 (formalism) を使用するべきかは指定しない。

【0108】

さらにまた、カラー・シグネチャ・ディスクリプタの記述フォーマットは、図 26 に示すようなものになり、カラー・シグネチャ・ディスクリプタは、下位要

素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがカラー・スペース (ColorSpace) であり、このヒストグラムがその中で計算される色彩空間を示し、実例値としてRGBが挙げられるカラー・スペース (ColorSpace) と、そのタイプがカラー (Color[n]) であり、この類似性を定義する色の集合であるn次元ベクトルを示すカラー (Colors) と、そのタイプがウエイト・ベクトル (WeightVector) であり、対応したColors型の色で記述されたイメージを表すWeightsのn次元ベクトルを示すウエイト (Weights) とを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (CHT) は、ディスクリプタコードを示している。

【0109】

このカラー・シグネチャ・ディスクリプタは、一般化された色ヒストグラムを表している。概念的に、ヒストグラムとは、あるイメージにおける色彩を何らかのカテゴリに割り当て、イメージ内におけるピクセルのそれぞれのカテゴリにおける出現頻度と見なすことができる。カラー・シグネチャにおいては、この方法を一般化し、任意のbin (色空間における最小分割空間) の設定を可能とする。ここで、各binは、その空間における代表的な色で表現される。

【0110】

例えば、あるカラーイメージを考える。このイメージで使用されているカラーを記述したい場合には、最初にカラーを領域に分割する。各領域は、分類項目となる。次にそれぞれのカラー領域へ分類されるイメージピクセルの相対的割合を決定する。これが特徴Weightingsとなる。

【0111】

つぎに、VBSに属するディスクリプタとして、シノプシス・ディスクリプタについて説明する。ここでは、ユーザに対して、ビデオデータをブラウジングするためのオーディオ・ビジュアル・サマリーを提供するディスクリプタについて記述する。このシノプシス・ディスクリプタの意図するところは、あるビデオデータの一部の内容を要約するために、視聴者に対して、視覚的、聴覚的、又はその両方により提供される何らかのAVエレメントを記述することである。

【0112】

このシノプシス・ディスクリプタの記述フォーマットは、図27に示すようなものになり、シノプシス・ディスクリプタは、スーパークラスとしてAVエレメント・ディスクリプタを有する。また、シノプシス・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがストリング (String) であり、シノプシスの種類を示し、実例値としてSLIDE

SHOWが挙げられる種別 (Kind) と、そのタイプがAVエレメントのリスト (list of AVElement) であり、そのシノプシスを構成するAVエレメントのリストを示すシーケンス (Sequence) と、そのタイプがブール (Boolean) であり、リスト内のエレメントの順序が重要かどうかを示すフラグであるIsOrderとを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (SYN) は、ディスクリプタコードを示している。

【0113】

このシノプシス・ディスクリプタは、AVエレメント中で描かれた動的出来事の要約を表す。これは、ユーザに対して概要を提供する際に表示すべきAVエレメントの順序を示している。

【0114】

一見すると、AVアブストラクトとシノプシスとは、同じような情報を表現していると思われる。しかしながら、そこには大きな違いがある。それは、シノプシスがユーザのためにAVエレメントを抽象化しているのに対し、AVアブストラクトにおいては、検索或いはAVへの索引付け (indexing) といったシステムのための抽象化を行っていることである。

【0115】

シノプシスの予想される第1の例としては、キーフレーム (Key frames) がある。シノプシス中の項目としては、AVエレメント中のアクションを描くために順序だてられた映像がある。

【0116】

また、第2の例としては、ダイジェスト (Digest) がある。映像及び音声の中の

項目は、AVセグメントである。例えば、ビデオデータからとられたクリップのシーケンスである。

【0117】

このシノプシス・ディスクリプタによって、ビデオデータの内容が概要として把握できるようになる。具体的には、シノプシス・ディスクリプタによって、後述するように、検出したシーンを、動的特徴のサンプリングにより検出した代表フレームで代表させ、ストーリーボードとして表示する。

【0118】

つぎに、構造ディスクリプタとして、音声フレーム・ディスクリプタと、映像フレーム・ディスクリプタと、ブラウジング・ノード・ディスクリプタと、ショット・ディスクリプタと、シーン・ディスクリプタと、トピック・ディスクリプタと、プログラム・ディスクリプタと、ハイライト・ディスクリプタとについて説明する。ビデオデータは、上述したように、3層の階層ストラクチャとして構築する。すなわち、ここでは、セグメント、シーン、トピックである。

【0119】

音声フレーム・ディスクリプタの記述フォーマットは、図28に示すようなものになり、音声フレーム・ディスクリプタは、スーパークラスとしてAVセグメント・ディスクリプタを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における(AF)は、ディスクリプタコードを示している。

【0120】

この音声フレーム・ディスクリプタは、単一の音声フレームに相当するAVセグメントを表す。

【0121】

また、映像フレーム・ディスクリプタの記述フォーマットは、図29に示すようなものになり、映像フレーム・ディスクリプタは、スーパークラスとしてAVセグメント・ディスクリプタを有する。また、映像フレーム・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがカラー・シグネチャ(ColorSignature)であり、そのフレームの色特徴を

示すカラー (Color) を有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (VF) は、ディスクリプタコードを示している。

【0122】

この映像フレーム・ディスクリプタは、AVセグメントの特別な場合であり、単一の映像フレームを表す。

【0123】

さらに、ブラウジング・ノード・ディスクリプタの記述フォーマットは、図30に示すようなものになり、ブラウジング・ノード・ディスクリプタは、スーパークラスとしてコンポジッド・セグメント・ディスクリプタを有する。また、ブラウジング・ノード・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがAVハイアラキの集合 (set of AVHierarchy) であり、あるAVセグメントの内容を、その下層にあるサブエレメントを用いてインデックス化するために使用される階層的分類の集合を示し、実例値としてビデオデータにおけるショットの分類が挙げられるインデックス (Index) と、そのタイプがシノプシス (Synopsis) であり、ビデオデータのある一部分における概要を示し、実例値としてキーフレームの並びが挙げられるシノプシス (Synopsis) と、そのタイプがA-Vグラフ (AVGraph) (任意) であり、ビデオデータのある一部分における構造を示すストラクチャ (Structure) とを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (BRO) は、ディスクリプタコードを示している。

【0124】

このブラウジング・ノード・ディスクリプタは、ブラウジングにて使用される様々なビデオストラクチャのユニットのための共通の抽象的クラスである。

【0125】

このブラウジング・ノード・ディスクリプタを用いることによって、ビデオデータの内容を具体的に記述する。また、ブラウジング・ノード・ディスクリプタを用いることによって、シノプシス・ディスクリプタやアクティビティ・レベル

・ディスクリプタの他、RDFノードで指定された意味的内容を記述することができる。

【0126】

さらにまた、ショット・ディスクリプタの記述フォーマットは、図31に示すようなものになり、ショット・ディスクリプタは、スーパークラスとしてブラウジング・ノード・ディスクリプタを有する。また、ショット・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがアクティビティ・レベル (ActivityLevel) であり、そのショットのアクティビティ・レベルを示し、実例値として例えば0.5が挙げられるアクティビティ (Activity) と、そのタイプがAVセグメントから (FromAVSegment) であるアブストラクト (Abstract) と、そのタイプがブラウジング・ノードから (FromBrowsingNode) であり、音声フレーム及び／又は映像フレームの並びから構成されるSynopsis型の概要を示すシノプシス (Synopsis) と、そのタイプがブラウジング・ノードから (FromBrowsingNode) であり、任意であるがショットを構成する音声フレーム及び／又は映像フレームの分類からなるショットへのインデックスを示すインデックス (Index) と、そのタイプがブラウジング・ノードから (FromBrowsingNode) であるストラクチャ (Structure) とを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (SHT) は、ディスクリプタコードを示している。また、アブストラクトは、ショットに対し、次のような抽象化を可能とする。まず第1としては、音声フレーム、映像フレーム、AVプロトタイプを要素とするテンポラル・シグネチャ (時間類似性) であり、第2としては、音声フレーム、映像フレームを要素とするシグネチャ (類似性) である。

【0127】

このショット・ディスクリプタは、オーディオ・ビジュアル・ユニットの最低レベル、すなわち、ショットを表す。

【0128】

ショット・ディスクリプタを用いることによって、ビデオデータのショットの内容を具体的に記述するものであり、ブラウジング・ノード・ディスクリプタを

継承してショットを記述するようにしたものである。具体的には、ショット・ディスクリプタを用いることによって、後述するように、シーンを検出する際に、そのクラスタリングの基になるデータであるショットを指定する。

【0129】

また、シーン・ディスクリプタの記述フォーマットは、図32に示すようなものになり、シーン・ディスクリプタは、スーパークラスとしてブラウジング・ノード・ディスクリプタを有する。また、シーン・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがストリング (String) であり、シーンの種類を示し、実例値としてDIALOGやCOMMERCIALやACTIONやANCHORが挙げられる種別 (kind) と、そのタイプがAVセグメントから (FromAVSegment) であるアブストラクト (Abstract) と、そのタイプがブラウジング・ノードから (FromBrowsingNode) であり、Shotsの並びから構成されるSynopsis型のデータからなるシーンへの可能な概要を示すシノプシス (Synopsis) と、そのタイプがブラウジング・ノードから (FromBrowsingNode) であり、シーンを構成するShotsの階層的分類からなるシーンへのインデックスを示すインデックス (Index) と、そのタイプがブラウジング・ノードから (FromBrowsingNode) であり、例えば、ノードが類似ショットのクラスタを表し、リンクがグループ間の推移を表すようなAVグラフを示すストラクチャ (Structure) とを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (SCN) は、ディスクリプタコードを示している。また、アブストラクトは、シーンに対しては、次のような抽象化が可能である。すなわち、ショット又はAVプロトタイプを要素として有する (時間) 類似性である。

【0130】

このシーン・ディスクリプタは、意味的に関連するショットのシーケンスを表す。

【0131】

このシーン・ディスクリプタを用いることによって、ビデオデータのシーンの内容を具体的に記述するものであり、ブラウジング・ノード・ディスクリプタを

継承してシーンを記述するようにしたものである。具体的には、シーン・ディスクリプタを用いることによって、後述するように、検出したシーンを指定する。

【0132】

さらに、トピック・ディスクリプタの記述フォーマットは、図33に示すようなものになり、トピック・ディスクリプタは、スーパークラスとしてブラウジング・ノード・ディスクリプタを有する。また、トピック・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがAVセグメントから (FromAVSegment) であるアブストラクト (Abstract) と、そのタイプがブラウジング・ノードから (FromBrowsingNode) であり、Scene sの並びから構成されるSynopsis型のデータからなるトピックへの可能な概要を示すシノプシス (Synopsis) と、そのタイプがブラウジング・ノードから (FromBrowsingNode) であり、トピックを構成するScenesの階層的分類からなるトピックへのインデックスを示すインデックス (Index) と、そのタイプがブラウジング・ノードから (FromBrowsingNode) であり、例えば、ノードが類似シーンのクラスタを表し、リンクがグループ間の推移を表すようなAVグラフを示すストラクチャ (Structure) とを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (TPC) は、ディスクリプタコードを示している。また、アブストラクトは、次のような抽象化が可能である。第1としては、シーン又はAVプロトタイプを要素として有する (時間) 類似性である。第2としては、ショットを値として有する (時間) 類似性である。第3としては、音声フレーム又は映像フレームを値として有する (時間) 類似性である。

【0133】

このトピック・ディスクリプタは、意味的に関連するシーンのシーケンスを表す。

【0134】

このトピック・ディスクリプタを用いることによって、ビデオデータのトピックの内容を具体的に記述するものであり、ブラウジング・ノード・ディスクリプタを継承してシーンを記述するようにしたものである。

【0135】

さらにまた、プログラム・ディスクリプタの記述フォーマットは、図34に示すようなものになり、プログラム・ディスクリプタは、スーパークラスとしてブラウジング・ノード・ディスクリプタを有する。また、プログラム・ディスクリプタは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがAVセグメントから (FromAVSegment) であるアブストラクト (Abstract) と、そのタイプがブラウジング・ノードから (FromBrowsingNode) であり、プログラムの概要として含まれるものには、Scenesの並びからなる概要と、Shotsの並びからなる概要と、このプログラムのハイライトシーンから構成される概要とがあることを示すシノプシス (Synopsis) と、そのタイプがブラウジング・ノードから (FromBrowsingNode) であるインデックス (Index) と、そのタイプがブラウジング・ノードから (FromBrowsingNode) であり、例えば、ノードが類似トピックのクラスタを表し、リンクがグループ間の推移を表すようなAVグラフを示すストラクチャ (Structure) とを有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (PRG) は、ディスクリプタコードを示している。また、アブストラクトは、次のような抽象化が可能である。第1としては、トピックを要素として有する (時間) 類似性である。第2としては、シーンを値として有する (時間) 類似性である。第3としては、ショットを値として有する (時間) 類似性である。さらに、プログラムへのインデックスは、全てのShotsへの階層的インデックスと全てのScenesへの階層的インデックスと全てのTopicsへの階層的インデックスといった様々なレベルにおけるインデックスより構成される。

【0136】

このプログラム・ディスクリプタは、トピックのシーケンスから構成されるプログラムを表す。

【0137】

また、ハイライト・ディスクリプタの記述フォーマットは、図35に示すようなものになり、ハイライト・ディスクリプタは、スーパークラスとしてブラウジング・ノード・ディスクリプタを有する。また、ハイライト・ディスクリプタは

、下位要素のディスクリプタ或いは数値の種類を示すフィーチャとして、そのタイプがウェイト (Weight) であり、そのハイライトへの推定興味レベルを示し、値が大きい程、興味レベルが高いことを示すインタレスト (Interest) を有する。なお、同図中におけるディスクリプタ名とディスクリプタコードとを示すディスクリプタの欄における (HI) は、ディスクリプタコードを示している。

【0138】

このハイライト・ディスクリプタは、放送番組で起こる何らかのハイライトとされる出来事を捉える。例えば、このハイライト・ディスクリプタは、野球の試合におけるホームランや映画における特に感動的な瞬間等をマークするのに使用される。各ハイライトには、そのハイライトが視聴者にどの程度面白いのか或いは感動的かを示す属性が付けられる。

【0139】

以上のようなディスクリプタの集合からなるディスクリプションスキームにおいては、様々なブラウジングノードにおけるindexに含まれる情報を用いることによって、階層構造を有するビデオデータのオーディオ・ビジュアル・インデックスを、様々な詳細レベルにおいて構築することが可能となる。例えば、Programにおいては、ビデオデータ内の全てのショットに対し、その類似性によって分類された階層構造が構築される。これを基にある詳細レベルの情報を表示したい場合には、階層構造を指定されたレベルで分け、その上の分類のみを用いればよい。その結果は、低層化された階層構造を有する分類となる。これら低層化されたグループに対し、それぞれのグループより、そのグループを代表するメンバーを選択することで、そのグループメンバーの映像及び音声の概要を表示することができる。

【0140】

また、ディスクリプションスキームにおいては、階層化ビデオストラクチャのそれぞれのレベルに関係付けられたsynopsisを用いることによって、そのビデオデータにおける階層化した映像音声の目次を取り出すことができる。例えば、それぞれのレベルが以下の表1で示すようなsynopsisを有する場合に、目次を表示したいときには、各ショットに関係付けられたキーフレームより、各レベルに対

するキーフレームを繰り返し取り出すことで映像音声の目次を取り出すことができる。

【0141】

【表1】

ビデオストラクチャのレベルに対するsynopsis

レベル	Synopsis
ショット	一連のキーフレーム
シーン	一連のキーショット
トピック	一連のキーシーン
プログラム	一連のキーシーン

【0142】

以下、ビデオデータからストラクチャを抽出する具体的な処理と、このような処理に対応して、上述したディスクリプションスキームとして定義される記述方法を用いて生成される抽出されたビデオストラクチャ等を記述する特徴データについて説明する。なお、ビデオデータからストラクチャを抽出する処理は、上述したデータ送受信処理システムにおける送信装置10及び受信装置20の両方において行うことができる。ここでは、送信装置10においてビデオデータからシーンストラクチャを抽出するとともに、それに対応して特徴データを生成し、ビデオデータとともに特徴データを受信装置20へと送信するものとして説明する。

【0143】

送信装置10は、ビデオデータにおける映像セグメント及び音声セグメントのフィーチャを用いてセグメント間の類似性を測定し、これらのセグメントをシーンにまとめてビデオストラクチャを自動的に抽出する。この際、時間的に近接して反復している類似したセグメントは、ほぼ同一シーンの一部であるため、送信装置10は、このようなセグメントを検出してグループ化していくことによって、シーンを検出する。このような送信装置10は、図36に概略を示すような一

連の処理を行うことによって、シーンを検出する。

【0144】

まず、送信装置10は、同図に示すように、ステップS1において、ビデオ分割を行う。すなわち、送信装置10は、入力されたビデオデータを映像セグメント又は音声セグメントのいずれか、或いは可能であればその両方に分割する。送信装置10は、適用するビデオ分割方法に特に前提要件を設けない。例えば、送信装置10は、“G. Ahanger and T.D.C. Little, A survey of technologies for parsing and indexing digital video, J. of Visual Communication and Image Representation 7:28-4, 1996”に記載されているような方法によりビデオ分割を行う。このようなビデオ分割の方法は、当該技術分野ではよく知られたものであり、送信装置10は、いかなるビデオ分割方法も適用できるものとする。

【0145】

次に、送信装置10は、ステップS2において、セグメントの特徴の抽出を行う。すなわち、送信装置10は、各セグメントについて1組の特徴を計算し、セグメントの特性を表す。送信装置10は、例えば、各セグメントの時間長や、カラーヒストグラムやテクスチャといった映像に関する特徴や、周波数解析結果、レベル、ピッチといった音声に関する特徴やアクティビティ測定結果等を適用可能な特徴として計算する。勿論、送信装置10は、適用可能な特徴としてこれらに限定されるものではない。

【0146】

さらに、送信装置10は、ステップS3において、特徴を用いたセグメントの類似性測定を行う。すなわち、送信装置10は、セグメント間の非類似性測定を行い、その測定基準により2つのセグメントがどの程度類似しているかを測定する。送信装置10は、先のステップS2において抽出した特徴を用いて、非類似性測定基準を計算する。

【0147】

そして、送信装置10は、ステップS4において、セグメントのグループ化を行う。すなわち、送信装置10は、先のステップS3において計算した非類似性測定基準と、先のステップS2において抽出した特徴とを用いて、時間的に近接

して類似したセグメントを繰り返しまとめ、これらのセグメントを重ねてグループに合わせる。送信装置10は、このようにして最終的に生成されたグループをシーンとして出力する。

【0148】

このような一連の処理を経ることによって、送信装置10は、ビデオデータからシーンを検出することができる。したがって、ユーザは、この結果を用いることによって、ビデオデータの内容を要約したり、ビデオデータ中の興味のあるポイントに迅速にアクセスしたりすることが可能となる。

【0149】

以下、同図に示した送信装置10における処理を各工程毎により詳細に説明していく。

【0150】

まず、ステップS1におけるビデオ分割について説明する。送信装置10は、入力されたビデオデータを映像セグメント又は音声セグメントのいずれか、或いは可能であればその両方に分割するが、このビデオデータにおけるセグメントの境界を自動的に検出するための技術は、多くのものがあり、送信装置10において、このビデオ分割方法に特別な前提要件を設けないことは上述した通りである。しかしながら、送信装置10において、後の工程によるシーン検出の精度は、本質的に、基礎となるビデオ分割の精度に依存する。なお、送信装置10におけるシーン検出は、ある程度ビデオ分割時のエラーを許容することができる。特に、送信装置10において、ビデオ分割は、セグメント検出が不十分である部分よりも、セグメント検出が過度である部分においてエラーを生じる方が好ましい。送信装置10は、類似したセグメントの検出が過度である結果である限り、たとえばセグメント検出が過度でなかったとしても、一般に、シーン検出の際に検出過度であるセグメントを同一シーンとしてまとめることができる。

【0151】

つぎに、ステップS2における特徴抽出について説明する。特徴は、セグメントの内容を表すとともに、異なるセグメント間の類似性を測定するためのデータを供給するセグメントの属性である。送信装置10は、各セグメントについて1

組の特徴を計算し、セグメントの特性を表す。送信装置 1 0 は、いかなる特徴の具体的詳細にも依存するものではない。送信装置 1 0 において用いて効果的であると考えられる特徴としては、例えば以下に示す映像に関する特徴、音声に関する特徴、映像音声ともにに関する特徴のようなものがある。送信装置 1 0 において適用可能であるこのような特徴は、非類似性測定を有していることが唯一の必要条件である。また、送信装置 1 0 は、実際には、特徴抽出と上述したビデオ分割とを効率化のために同時に行うことがあり、以下に説明する特徴は、このような処理を可能にするものである。

【0 1 5 2】

特徴としては、まず映像に関する特徴が挙げられる。セグメントを構成する映像（画像）は、そのセグメントの描写内容の大部分を表している。そのため、映像セグメントの類似性は、映像そのものの類似性に変えることができる場合が多い。したがって、映像に関する特徴は、送信装置 1 0 で用いることができる重要な特徴の 1 種である。この映像に関する特徴は、動的な情報よりも静的な情報を表すため、送信装置 1 0 は、後述するような方法によって、動的情報を得るように映像セグメント内の映像に関する特徴を抽出する。

【0 1 5 3】

映像に関する特徴として既知のものは、多数存在するが、送信装置 1 0 は、映像に関する特徴として色特徴（ヒストグラム）及び映像相関を用いる。

【0 1 5 4】

送信装置 1 0 において、映像における色は、2 つの映像が類似しているかを判断する際の重要な材料となる。カラーヒストグラムを用いて映像の類似性を判断することは、例えば “G. Ahanger and T.D.C. Little, A survey of technologies for parsing and indexing digital video, J. of Visual Communication and Image Representation 7:28-4, 1996” に記載されているように、よく知られている。カラーヒストグラムは、例えば HSV や RGB 等の 3 次元色空間を n 個の領域に分割し、各領域にある映像における画素の相対的割合を計算したものである。そして、得られた情報からは、 n 次元ベクトルが与えられる。圧縮されたビデオデータについては、例えば U.S. Patent #5,708,767 号公報に記載されてい

るように、カラーヒストグラムは、圧縮データから直接抽出される。

【0155】

送信装置10は、セグメントを構成する映像におけるもともとのYUV色空間を、色チャンネル当たり2ビットでサンプルし、長さ $2^{2 \cdot 3} = 64$ のヒストグラムを得る。

【0156】

このようなヒストグラムは、映像の全体的な色調を表すが、空間的情報を欠いている。そこで、送信装置10においては、映像相関を映像に関する特徴の1つとして計算する。送信装置10におけるシーン検出に関しては、インターリーブされたセグメントは、シーンストラクチャの有力な指標を与える。送信装置10は、映像相関を計算するために、映像を $M \times N$ の大きさのグレイスケール映像にサブサンプルする。ここで、 M と N は、両方とも小さい値であり、例えば 8×8 である。小さいグレイスケール映像は、長さ MN の特徴として解釈される。

【0157】

さらに上述した映像に関する特徴の他の特徴としては、音声に関する特徴が挙げられる。音声に関する特徴は、音声セグメントの内容を表すのに用いることができるものである。送信装置10は、音声に関する特徴として周波数解析、ピッチ、レベルを用いる。これらの音声に関する特徴は、種々の文献により知られているものである。

【0158】

まず、送信装置10は、フーリエ変換等の周波数解析を行うことによって、単一の音声フレームにおける周波数情報の分布を決定することができる。送信装置10は、例えば、1つの音声セグメントにわたる周波数情報の分布を表すために、FFT (Fast Fourier Transform; 高速フーリエ変換) 成分、周波数ヒストグラム、パワースペクトル、その他の特徴を用いることができる。

【0159】

また、送信装置10は、平均ピッチや最大ピッチといったピッチや、平均ラウドネスや最大ラウドネスといったレベルを音声セグメントを表す有効な音声に関する特徴として用いることができる。

【0160】

さらに他の特徴としては、映像音声ともにに関する特徴が挙げられる。映像音声ともにに関する特徴は、特に映像に関する特徴でもなく音声に関する特徴でもないが、送信装置10において、シーン内のセグメントの内容を表すのに有用な情報を与えるものである。送信装置10は、この映像音声ともにに関する特徴として、セグメント長とアクティビティとを用いる。

【0161】

送信装置10は、映像音声ともにに関する特徴として、セグメント長を用いる。このセグメント長は、セグメントにおける時間長である。送信装置10において、シーンは、よくその変化のリズムの特性を有し、シーン内のセグメント長の傾向と同様の傾向をとる。例えば、迅速に連なった短いセグメントは、コマーシャルを表す。ところが、会話場面におけるセグメントは、より長くなるが、その長さの中でお互いに類似している。送信装置10は、このような特性を有するセグメント長を映像音声ともにに関する特徴として用いることができる。

【0162】

また、送信装置10は、映像音声ともにに関する特徴として、アクティビティを用いる。アクティビティは、セグメントの内容がどの程度動的或いは静的であるように感じられるかを表す。例えば、アクティビティは、視覚的に動的である場合には、カメラが対象物に沿って迅速に移動する度合若しくは撮影されている物が迅速に変化する度合を表す。

【0163】

このアクティビティは、カラーヒストグラムのような特徴の平均フレーム間非類似性を測定することにより間接的に計算される。ここで、フレーム*i*とフレーム*j*との間で測定された特徴*F*に対する非類似性測定基準を $d_F(i, j)$ と定義すると、映像アクティビティ V_F は、次式(1)のように定義される。

【0164】

【数1】

$$V_F = \frac{\sum_{i=b}^{f-1} d_F(i, i+1)}{f-b+1} \quad \dots (1)$$

【0165】

この式(1)において、 b と f は、それぞれ、1セグメントにおける最初と最後のフレームのフレーム番号である。送信装置10は、具体的には、例えば上述したヒストグラムを用いて、映像アクティビティ V_F を計算する。

【0166】

ところで、上述した映像に関する特徴を始めとする特徴は、基本的にセグメントの静的情報を表すものであることは上述した通りであるが、セグメントの特徴を正確に表すためには、動的情報を表す必要がある。そこで、送信装置10は、以下に示すような特徴のサンプリング方法により動的情報を表す。

【0167】

送信装置10は、例えば図37に示すように、1セグメント内の異なる時点から1以上の静的な特徴値の組を抽出する。このとき、送信装置10は、特徴の抽出数を、忠実度の最大化と冗長度の最小化との釣り合いをとることにより決定する。例えば、セグメント内の1画像がキーフレームとして指定された場合には、そのキーフレームについて計算されたヒストグラムが抽出された特徴となる。

【0168】

送信装置10は、サンプリング方法を用いて、セグメントにおいて、ある特徴がとる全ての値のうち、どのサンプルを選択するかを決定する。送信装置10は、優れたサンプリング方法を必要とする。

【0169】

ここで、あるサンプルが常に所定の時点、例えばセグメント内の最後の時点においてとられる場合を考える。この場合、黒にフェードする任意の2つのセグメントについては、サンプルが同一の黒いフレームとなるため、同一の特徴値が得

られる結果になる恐れがある。すなわち、これらのセグメントの映像内容がいかなるものであれ、サンプルした2つのフレームは、極めて類似していると判断されてしまう。このような問題は、サンプルが良好な代表値でないために発生するものである。

【0170】

そこで、送信装置10は、このように固定点で特徴を抽出するのではなく、統計的な代表値を抽出する。ここでは、一般的な特徴のサンプリング方法を2つの場合、すなわち、第1の場合として、特徴を実数の n 次元ベクトルとして表すことができる場合と、第2の場合として、非類似性測定基準しか利用できない場合とについて説明する。なお、第1の場合は、ヒストグラムやパワースペクトル等、最もよく知られている映像に関する特徴及び音声に関する特徴を含むものである。

【0171】

第1の場合には、サンプル数は、事前に k と決められており、送信装置10は、“L. Kaufman and P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, John-Wiley and sons, 1990”に記載されてよく知られている k 平均値クラスタリング法 (k -means-clustering method) を用いて、セグメント全体についての特徴値を k 個の異なるグループに自動的に分割する。そして、送信装置10は、 k 個の各グループから1サンプルを選択する。すなわち、送信装置10は、グループのセントロイド (平均ベクトル) 又はグループのセントロイドに近いサンプルを選択する。送信装置10は、この処理を短時間で行うことができ、サンプル数において直線的な時間しか要さない。

【0172】

一方、第2の場合には、送信装置10は、“L. Kaufman and P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, John-Wiley and sons, 1990”に記載されている k -メドイドアルゴリズム法 (k -medoids algorithm method) を用いて、 k 個のグループを形成する。そして、送信装置10は、各グループ毎に、サンプル値として、上述したグループのセントロイドに類似するグループのメドイドを用いる。

【0173】

このようにして、送信装置10は、静的な特徴を抽出し、動的情報を表すことができる。

【0174】

以上のように、送信装置10は、種々の特徴を抽出することができる。これらの各特徴は、一般に、単一ではセグメントの内容を表すのに不十分であることが多い。そこで、送信装置10は、これらの各種特徴を組み合わせることで、互いに補完し合う特徴の組を選択することができる。例えば、送信装置10は、上述したカラーヒストグラムと映像相関とを組み合わせることによって、各特徴が有する情報よりも多くの情報を得ることができる。

【0175】

つぎに、図36中ステップS3における特徴を用いたセグメントの類似性測定について説明する。送信装置10は、2つの特徴について、その2つの特徴値がどの程度非類似であるかを測定する実数値を計算する関数である非類似性測定基準を用いて、セグメントの類似性測定を行う。この非類似性測定基準は、その値が小さい場合は2つの特徴が類似していることを示し、値が大きい場合は非類似であることを示す。ここでは、特徴Fに関する2つのセグメント S_1 、 S_2 の非類似性を計算する関数を非類似性測定基準 $d_F(S_1, S_2)$ と定義する。このような関数は、以下の式(2)に示す特性に合致するものでなければならない。

【0176】

【数2】

$$\begin{aligned} d_F(S_1, S_2) &= 0 && (S_1 = S_2 \text{ のとき}) \\ d_F(S_1, S_2) &\geq 0 && (\text{全ての } S_1, S_2 \text{ について}) \\ d_F(S_1, S_2) &= d_F(S_2, S_1) && (\text{全ての } S_1, S_2 \text{ について}) \end{aligned} \quad \dots (2)$$

【0177】

なお、適切な非類似性測定基準が、特定の特徴に依存することがあるが、“G. Ahanger and T.D.C. Little, A survey of technologies for parsing and ind

exing digital video, J. of Visual Communication and Image Representation 7:28-4, 1996” や “L. Kaufman and P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, John-Wiley and sons, 1990” に記載されているように、多くの一般的な非類似性測定基準は、 n 次元空間における点として表される特徴についての類似性を測定するのに有効であるとして知られている。これらは、ユークリッド距離、内積、 $L1$ 距離等である。特に、 $L1$ 距離が、ヒストグラムや映像相関といった特徴を含む種々の特徴に関して、有効に作用することを見出したことから、送信装置 10 は、2つの n 次元ベクトル A , B 間の $L1$ 距離を次式 (3) のように求める。

【0178】

【数3】

$$d_{L1}(A,B) = \sum_{i=1}^n |A_i - B_i| \quad \dots (3)$$

【0179】

ここで、下付文字は、 n 次元ベクトルの i 番目の要素を示すものである。

【0180】

また、送信装置 10 は、上述したように、時間とともに変化する特徴に関して、セグメントにおける様々な時点で特徴値を抽出する。そして、送信装置 10 は、2つの抽出された特徴の間の類似性を決定するために、抽出された特徴の非類似性を、その基礎となる特徴の非類似性測定基準により定義する。多くの場合、非類似性は、2つの抽出された特徴のそれぞれから選択された最も非類似性のない特徴の対の非類似性値を用いて決定されるのが最良である。すなわち、送信装置 10 は、最小非類似性を決定する。この場合、2つの抽出された特徴 SF_1 , SF_2 の間のサンプルされた非類似性測定基準は、次式 (4) のように定義される。

【0181】

【数4】

$$d_s(SF_1, SF_2) = \min_{F_1 \in SF_1, F_2 \in SF_2} d_F(F_1, F_2) \quad \dots (4)$$

【0182】

上式(4)における関数 $d_F(F_1, F_2)$ は、その基礎となる抽出された特徴 F についての非類似性測定基準を示す。また、非類似性の値の最小をとる代わりに、最大又は平均をとる方がより適切である場合もある。

【0183】

さらに、送信装置10は、多くの場合、類似性を決定するために、同一セグメントについての多数の特徴から情報を組み合わせる必要がある。この1つの方法として、送信装置10は、種々の特徴ベクトル非類似性関数の重み付きの組み合わせを計算する。すなわち、送信装置10は、 k 個の特徴 F_1, F_2, \dots, F_k が存在する場合、次式(5)に表される組み合わせた特徴に関する非類似性測定基準 $d_F(S_1, S_2)$ を用いる。

【0184】

【数5】

$$d_F(S_1, S_2) = \sum_{i=1}^k w_i d_{Fi}(S_1, S_2) \quad \dots (5)$$

【0185】

ここで、 $\{w_i\}$ は、重みと $\sum_i w_i = 1$ との組である。

【0186】

以上のように、送信装置10は、図36中ステップS2において抽出した特徴を用いて非類似性測定基準を計算し、セグメントの類似性を測定することができる。

【0187】

つぎに、図36中ステップS4におけるセグメントのグループ化について説明する。送信装置10は、非類似性測定基準と抽出した特徴とを用いて、時間的に近接して類似したセグメントを繰り返しまとめ、これらのセグメントを重ねてグループに合わせ、最終的に生成されたグループをシーンとして出力する。

【0188】

送信装置10は、セグメントをグループ化してシーンを検出する際に、2つの基本的な処理を行う。送信装置10は、まず第1の処理として、互いに時間的に近接して類似したセグメントのグループを検出する。この処理により得られるグループは、ほとんどが同一シーンの一部となるものである。また、送信装置10は、セグメントが時間的に重複しているため、第2の処理として、互いに時間が重複して類似したシーンをまとめる。送信装置10は、このような処理を各セグメントから開始し、反復して繰り返す。そして、送信装置10は、徐々にセグメントのグループを大きく構築していき、最終的に生成したグループをシーンの組として出力する。

【0189】

このような処理において、送信装置10は、その処理動作を制御するために2つの制約を用いる。

【0190】

すなわち、送信装置10は、第1の制約として、2つのセグメントが、どの程度類似している場合に同一のシーンのものであるのに十分類似しているかと考えられるかを決定する非類似性閾値 δ_{sim} を用いる。例えば、図38に示すように、送信装置10は、あるセグメントに対して一方のセグメントが類似性領域に属するか非類似性領域に属するかを判断する。

【0191】

なお、送信装置10は、非類似性閾値 δ_{sim} をユーザにより設定するようにしてもよく、自動的に決定してもよい。

【0192】

また、送信装置10は、第2の制約として、2つのセグメントが、どの程度広

く分割されている場合に同一のシーンのものであるかとなお考えられ得るかを決定する時間閾値Tを用いる。例えば、図39に示すように、送信装置10は、時間閾値Tの範囲内で互いに近接して続いている類似した2つのセグメントA、Bを同一シーンにまとめるが、時間的に大きく離れていて時間閾値Tの範囲外である2つのセグメントB、Cをまとめることはない。このように、送信装置10は、この時間閾値Tによる時間制約があるために、互いに類似しているものの大きく離れているセグメントを同一シーンにまとめてしまうエラーを発生することがない。

【0193】

送信装置10は、類似セグメントのグループを求めるために、ここでは、“L. Kaufman and P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, John-Wiley and sons, 1990”に記載されている階層的クラスタ分割方法 (hierarchical clustering method) を適合させて用いることにする。このアルゴリズムにおいては、2つのクラスタ C_1 、 C_2 間の非類似性測定基準 $d_C(C_1, C_2)$ について、次式(6)に示すように、含まれる2つの要素間の最小非類似性として定義する。

【0194】

【数6】

$$d_C(C_1, C_2) \equiv \min_{S_1 \in C_1, S_2 \in C_2} dist_S(S_1, S_2) \quad \dots (6)$$

【0195】

なお、送信装置10においては、上式(6)で示される最小関数を最大又は平均関数に容易に置換することができる。

【0196】

まず、送信装置10は、図40に示すように、ステップS11において、変数Nをセグメントの数に初期化する。この変数Nは、常に現在検出されているグループの数を示すものである。

【0197】

次に、送信装置10は、ステップS12において、クラスタの組を生成する。送信装置10は、初期時には、N個の各セグメントを異なるクラスタとする。そのため、初期時には、N個のクラスタが存在することになる。各クラスタは、 C^{start} と C^{end} とによって表されるその開始時と終了時とを示す特性を有する。クラスタの組の要素は、 C^{start} から順番に整頓される。

【0198】

さらに、送信装置10は、ステップS13において、変数 t を1に初期化し、ステップS14において、変数 t が時間閾値 T よりも大きいかな否かを判別する。ここで、送信装置10は、変数 t が時間閾値 T よりも大きい場合には、ステップS23へと処理を移行し、変数 t が時間閾値 T よりも小さい場合には、ステップS15へと処理を移行する。ただし、ここでは、変数 t が1であるため、送信装置10は、ステップS15へと処理を移行する。

【0199】

送信装置10は、ステップS15において、非類似性測定基準 d_C を計算し、N個のクラスタの中から最も類似した2つのクラスタを検出する。ただし、ここでは、変数 t が1であるため、送信装置10は、隣接したクラスタ間の非類似性測定基準 d_C を計算し、その中から最も類似したクラスタの対を検出する。

【0200】

このような最も類似した2つのクラスタを検出する方法としては、対象となる全てのクラスタの対を走査することが考えられる。送信装置10は、時間閾値 T による時間制約を有するため、対象とするクラスタの対の数を限定することができる。そのため、送信装置10は、 t 個のセグメントにより分割されているクラスタのみを走査すればよい。クラスタの組は、時間順に整頓されているため、送信装置10は、あるクラスタに対してその前方及び後方へと走査していき、 t 個のセグメントよりも離れたクラスタを走査すれば、このクラスタ以降の全てのセグメントを対象外として走査処理を終了することができる。

【0201】

ここで、検出された2つのクラスタをそれぞれ C_i 、 C_j と定義し、これらのク

クラスタ C_i , C_j の間の非類似性の値を d_{ij} と定義する。

【0202】

送信装置 10 は、ステップ S16 において、非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも大きいかなんかを判別する。ここで、送信装置 10 は、非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも大きい場合には、ステップ S21 へと処理を移行し、非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも小さい場合には、ステップ S17 へと処理を移行する。ここでは、非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも小さいものとする。

【0203】

送信装置 10 は、ステップ S17 において、クラスタ C_j をクラスタ C_i に合わせる。すなわち、送信装置 10 は、クラスタ C_j の要素の全てをクラスタ C_i に加える。

【0204】

また、送信装置 10 は、ステップ S18 において、クラスタ C_j をクラスタの組から除去する。なお、2つのクラスタ C_i , C_j を合わせるにより開始時 C_i^{start} の値が変化した場合には、送信装置 10 は、開始時による順序を維持するために、クラスタの組の要素を再び並べ替える。

【0205】

さらに、送信装置 10 は、ステップ S19 において、変数 N から 1 を減じる。

【0206】

そして、送信装置 10 は、ステップ S20 において、変数 N が 1 であるかなんかを判別する。ここで、送信装置 10 は、変数 N が 1 である場合には、ステップ S23 へと処理を移行し、変数 N が 1 でない場合には、ステップ S15 へと処理を移行する。ここでは、変数 N が 1 でないものとする。

【0207】

すると、送信装置 10 は、ステップ S15 において、再び非類似性測定基準 d_c を計算し、 $N-1$ 個のクラスタの中から最も類似した 2つのクラスタを検出する。ここでも、変数 t が 1 であるため、送信装置 10 は、隣接したクラスタ間の非類似性測定基準 d_c を計算し、その中から最も類似したクラスタの対を検出す

る。

【0208】

また、送信装置10は、ステップS16において、非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも大きいか否かを判別する。ここでも、非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも小さいものとする。

【0209】

そして、送信装置10は、ステップS17乃至ステップS20の処理を行う。

【0210】

送信装置10は、このような処理を繰り返し、変数Nが減算されていった結果、ステップS20において、変数Nが1であると判別した場合には、ステップS23において、単一のセグメントのみを含むクラスタを合わせる。すなわち、送信装置10は、この場合は、全てのセグメントが1つのクラスタにまとめられているため、この処理を行う必要はなく、一連の処理を終了する。

【0211】

さて、送信装置10は、ステップS16において、非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも大きいと判別した場合には、ステップS21へと処理を移行するが、この場合には、ステップS21において、時間的に重複しているクラスタを繰り返し合わせる。すなわち、 C_i の時間間隔 $[C_i^{start}, C_i^{end}]$ が、 C_j の時間間隔 $[C_j^{start}, C_j^{end}]$ と相交している場合には、2つのクラスタ C_i と C_j は、重複しているため、送信装置10は、クラスタをその組の開始時により整頓すると、重複しているクラスタを検出し、1つに合わせることができる。

【0212】

そして、送信装置10は、ステップS22において、変数 t に1を加算して $t=2$ とし、ステップS14へと処理を移行して変数 t が時間閾値 T よりも大きいか否かを判別する。ここでも、変数 t が時間閾値 T よりも小さいものとし、送信装置10は、ステップS15へと処理を移行するものとする。

【0213】

送信装置10は、ステップS15において、非類似性測定基準 d_c を計算し、現在存在する複数のクラスタの中から最も類似した2つのクラスタを検出する。

ただし、ここでは、変数 t が 2 であるため、送信装置 10 は、1 つおきに離れているクラスタ間の非類似性測定基準 d_c を計算し、その中から最も類似したクラスタの対を検出する。

【0214】

そして、送信装置 10 は、ステップ S16 において、1 つおきに離れているクラスタ C_i 、 C_j の非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも大きいかなんかを判別する。ここでも、非類似性値 d_{ij} が非類似性閾値 δ_{sim} よりも大きいものとし、送信装置 10 は、ステップ S21 及びステップ S22 の処理を経て、変数 t に 1 を加算して $t = 3$ として再びステップ S14 以降の処理へと移行する。勿論、送信装置 10 は、変数 t が 3 のときには、ステップ S15 において、2 つおきに離れているクラスタ間の非類似性測定基準 d_c を計算し、その中から最も類似したクラスタの対を検出する。

【0215】

送信装置 10 は、このような処理を繰り返し、変数 t が加算されていった結果、ステップ S14 において、変数 t が時間閾値 T よりも大きいと判別すると、ステップ S23 へと処理を移行し、単一のセグメントのみを含むクラスタを合わせる。すなわち、送信装置 10 は、孤立しているクラスタを単一の要素のみを含むクラスタとみなし、このような一連のクラスタが存在している場合には、これらのクラスタを単一のクラスタに合わせる。この工程は、単一のシーンと類似性関連を有さないセグメントをまとめるものである。なお、送信装置 10 は、必ずしもこの工程を行う必要はない。

【0216】

そして、送信装置 10 は、ステップ S22 の処理を経て一連の処理を終了する。

【0217】

このような一連の処理によって、送信装置 10 は、複数のクラスタをまとめていき、シーンを生成することができる。

【0218】

送信装置 10 は、シーンを検出する際に、セグメントが同一グループであるか

を決定するために、単一の非類似性測定基準をもちいるばかりではなく、重み付け関数を用いて、異種の特徴に関する多様な非類似性測定基準を組み合わせることができることは上述した通りである。送信装置 10 において、このような特徴の重み付けは、試行錯誤の末得られるものであり、各特徴が質的に異なるタイプのものである場合には、通常、適切な重み付けを行うことは困難であるが、送信装置 10 は、例えば、カラーヒストグラムとテクスチャとを組み合わせる場合、各特徴に関してシーンを検出し、検出された各シーンストラクチャを単一のシーンストラクチャに組み合わせることができる。ここで、各特徴に関してシーンを検出した結果をシーン層と称することにする。例えば、特徴としてカラーヒストグラムとセグメント長とを用いる場合、送信装置 10 は、これらの特徴に対してシーンを検出し、カラーヒストグラムについてのシーン層と、セグメント長についてのシーン層とを得て、これらのシーン層を単一のシーンストラクチャに組み合わせることができる。

【0219】

さらに、一般に、映像領域と音声領域とからの情報を組み合わせることはできないが、送信装置 10 は、質的に異なるタイプの特徴を組み合わせる場合と同様な方法により、映像領域と音声領域とから得られるシーン層を単一のシーンストラクチャに組み合わせることができる。

【0220】

このような処理について説明する。ここでは、それぞれが類似性の 1 つの基準を表す k 個の特徴 F_1, F_2, \dots, F_k があるものとし、各特徴 F_i に対応して、非類似性測定基準 d_F^i と、非類似性閾値 δ_{sim}^i と、時間閾値 T^i とがあるものとする。送信装置 10 は、これらの各特徴 F_i に対する非類似性測定基準 d_F^i と、非類似性閾値 δ_{sim}^i と、時間閾値 T^i とを用いてシーン層の組 $X_i = \{X_i^j\}$ を検出する。また、送信装置 10 は、映像情報と音声情報とに対して分割的にシーン層を検出し、映像情報と音声情報とに関する 2 つの独立したシーン層 $X_i = \{X_i^j\}$ ($i = 1, 2$) を生成するものとする。

【0221】

送信装置 10 は、異なるシーン層を単一のシーンストラクチャに組み合わせる

ため、シーン境界の組み合わせ方を決定する必要がある。このシーン境界は、お互いにそろっている保証はないものである。ここで、各シーン層に関して、シーン境界を示す一連の時間で表される境界点 $t_{i1}, t_{i2}, \dots, t_{i|X_i|}$ があるものとする。まず、送信装置 10 は、種々のシーン層を単一のグループに組み合わせるために、最初にあるシーン層を整列に関する基礎とするために選択する。そして、送信装置 10 は、最終的に組み合わせ生成するシーンストラクチャにおけるシーン境界かどうかを各境界点 $t_{i1}, t_{i2}, \dots, t_{i|X_i|}$ に関して決定する。

【0222】

ここで、シーン層の組の番号である各 i に関して、 $B_i(t)$ を、 i 番目のシーン層 X_i において、ある時間 t で「近く」にシーン境界があるかどうかを示すブール関数とする。この「近く」の意味は、変化し、例えば、映像情報と音声情報とをそろえている場合には、0.5 秒であったりする。

【0223】

送信装置 10 は、各境界点 $t_j = t_{ij}, j = 1, \dots, |X_i|$ に関して、 $1 = 1, \dots, k$ のそれぞれについて、関数 $B_1(t_j)$ の結果を計算する。この結果は、各独立したシーン層に関して、シーン層 X_i において時間 t の近くにシーン境界があるかどうかを示している。そして、送信装置 10 は、決定関数として、組み合わせられた決定において時間 t がシーン境界であるかどうかを決定するための $B_i(t_j)$ の値を用いる。

【0224】

このような決定関数の 1 つの単純な例は、 $B_i(t_j)$ が定数 m 以上の 1 に等しい値の総数を計算することであり、時間の点 t は、最後のシーンストラクチャにおけるシーン境界を明示する。特に、 $m = 1$ の場合、これは、「oring」境界点と同義のものである。 $m = k$ の場合、全ての境界点をそろえる要求と同義である。

【0225】

このようにして、送信装置 10 は、異なるシーン層を単一のシーンストラクチャに組み合わせることができる。

【0226】

このような送信装置10における処理は、上述したディスクリプタを用いて記述することができる。

【0227】

図36中ステップS2における特徴抽出工程を記述するために用いられるディスクリプタとしては、例えば、アクティビティ・レベル・ディスクリプタ、シグネチャ・ディスクリプタ、テンポラル・シグネチャ・ディスクリプタ、シノプシス・ディスクリプタ等がある。

【0228】

アクティビティ・レベル・ディスクリプタは、特徴抽出工程における映像音声ともに関する特徴としてのアクティビティを記述するのに用いられる。また、シグネチャ・ディスクリプタは、特徴抽出工程における動的情報を表すための特徴のサンプリング方法を記述するのに用いられる。さらに、テンポラル・シグネチャ・ディスクリプタは、特徴抽出工程における動的情報を表すための最適な特徴を探し出す過程にて用いることができる。さらにまた、シノプシス・ディスクリプタは、同図中ステップS4において検出したシーンを、特徴抽出工程における動的特徴の抽出により検出した代表フレームで代表させる際に用いられる。

【0229】

また、同図中ステップS3における特徴を用いたセグメントの類似性測定工程を記述するために用いられるディスクリプタとしては、例えば、ウェイト・ベクトル・ディスクリプタ等がある。

【0230】

ウェイト・ベクトル・ディスクリプタは、類似性測定工程において多数の特徴から情報を組み合わせる際の重み付けの際に用いられる。

【0231】

さらに、同図中ステップS4におけるセグメントのグループ化工程を記述するために用いられるディスクリプタとしては、例えば、AVプロトタイプ・ディスクリプタ、クラスタ・ディスクリプタ、コンポジッド・セグメント・ディスクリプタ、AVグラフ・ディスクリプタ、ショット・ディスクリプタ、シーン・ディ

スクリプタ等がある。

【0232】

AVプロトタイプ・ディスクリプタやクラスタ・ディスクリプタは、グループ化工程におけるクラスタリング処理を記述するのに用いられる。また、コンポジット・セグメント・ディスクリプタは、グループ化工程においてセグメントやシーンをクラスタリングした結果を記述するのに用いられる。さらに、AVグラフ・ディスクリプタは、グループ化工程において検出したシーンの内容を、そのシーンの反復構造として記述するのに用いられる。さらにまた、ショット・ディスクリプタは、グループ化工程においてシーンを検出する際の基となるショットを指定するのに用いられ、シーン・ディスクリプタは、グループ化工程において検出したシーンを指定するのに用いられる。

【0233】

したがって、あるビデオデータにおけるシーンストラクチャを記述するディスクリプションスキームは、複数のディスクリプタの集合により表される。この具体例として、ある1つのシーンを記述するディスクリプションスキーム及び特徴データを図4-1及び図4-2に示す。ここで対象とするシーンは、開始位置がフレーム番号“12547”であり、終了位置がフレーム番号“12830”であるものとする。なお、図4-2中でセミコロン(;)の後の記載は、この行において参照すべきディスクリプタを示している。

【0234】

まず、特徴データは、図4-2における1行目でシーン記述を示すディスクリプタの始まりを示す。このシーンに対応するディスクリプションスキーム上でのノードは、図4-1中の頂点におけるノードである。

【0235】

次に、特徴データは、図4-2における2、3行目で、それぞれ、シーンの開始位置 (startPoint) 及び終了位置 (endPoint) を、時間やフレーム番号等により指定する。ディスクリプションスキーム上でのこの情報は、図4-1中においてスタートタイム (startTime) 及びエンドタイム (endTime) により表される。

【0236】

また、特徴データは、図42における4行目でビデオコンテンツの中におけるこのシーンが存在する場所を指定する。

【0237】

さらに、特徴データは、同図における5行目でシノプシス (synopsis) ・ディスクリプタの始まりを示し、ユーザにブラウジングさせるためのキーフレーム (KEY_FRAME) を指定する。

【0238】

次に、特徴データは、同図における6行目でシーケンス (sequence) ・ディスクリプタの始まりを示し、以下にAVエレメント (AVElement) のシーケンスがあることを示す。ここでは、AVエレメントは、キーフレームである。

【0239】

また、特徴データは、同図における7、8、9行目で、それぞれ、1、2、3番目のシノプシスのキーフレームをKEY_FRAME 1、KEY_FRAME 2、KEY_FRAME 3と指定する。ただし、ここでは、KEY_FRAME 1、KEY_FRAME 2、KEY_FRAME 3が具体的にどのフレームであるかという指定を行っていないが、実際には、どこかで指定する必要がある。

【0240】

さらに、特徴データは、同図における10、11行目で、それぞれ、シーケンス・ディスクリプタの終わり及びを示すシノプシス・ディスクリプタの終わりを示す。

【0241】

次に、特徴データは、同図における12行目でアブストラクト (abstract) ・ディスクリプタの始まりを示し、ビデオコンテンツの内容を代表する部分の指定を行う。

【0242】

また、特徴データは、同図における13行目でシグネチャ (signature) ・ディスクリプタの始まりを示し、クラスタリング等に用いるためのシーンの部分を

指定する。ここでは、ショット・シグネチャ (SHOT_SIGNATURE) で示されることが宣言されている。

【0243】

さらに、特徴データは、同図における14行目でシグネチャのメンバー (members) の指定開始を示す。

【0244】

そして、特徴データは、同図における15、16行目でシグネチャのメンバーの1つ目がshot1、2つ目がshot2、3つ目がshot3であることを示す。このshot1及びshot2は、22行目及び32行目に定義されている。

【0245】

また、特徴データは、同図における17行目でシグネチャのメンバーの指定終了を示す。

【0246】

次に、特徴データは、同図における18行目でシグネチャのメンバーに対する重み付けを指定する。ここでは、shot1には0.1、shot2及びshot3には0.45の重要度があることを示している。

【0247】

そして、特徴データは、同図における19、20行目で、それぞれ、シグネチャ・ディスクリプタ及びアブストラクト・ディスクリプタの終わりを示す。

【0248】

次に、特徴データは、同図における21行目でコンポーネント (components) ・ディスクリプタの始まりを示し、このシーンの構成要素を指定する。このコンポーネントに対応するディスクリプションスキーム上でのノードは、図41中のシーンを表すノードの一段だけ下位のノードである。

【0249】

また、特徴データは、図42における22行目でショット (shot) ・ディスクリプタの始まりを示し、第1番目のこのシーンの構成要素がshot1であることを指定し、その内容を以下に記述することを宣言する。

【0250】

さらに、特徴データは、同図における23、24行目で、それぞれ、shot 1の開始位置及び終了位置を、時間やフレーム番号等により指定する。ここでは、開始位置がフレーム番号“12547”であり、終了位置がフレーム番号“12616”であることを示している。ディスクリプションスキーム上でのこの情報は、図41中のコンポーネントを表すノードの一段だけ下位のノードにおいてスタートタイム (startTime) 及びエンドタイム (endTime) により表される。

【0251】

次に、特徴データは、図42における25行目でアブストラクト・ディスクリプタの始まりを示し、shot 1の内容を代表する部分の指定を行う。

【0252】

また、特徴データは、同図における26行目でシグネチャ・ディスクリプタの始まりを示し、クラスタリング等に用いるためのshot 1の部分を指定する。ここでは、フレーム・シグネチャ (FRAME_SIGNATURE) で示されることが宣言されている。

【0253】

さらに、特徴データは、同図における27行目でシグネチャのメンバーが1つだけであることを示し、そのフレーム番号を指定する。ここでは、フレーム番号が“12590”であることが記述されている。

【0254】

また、特徴データは、同図における28行目でシグネチャのメンバーに対する重み付けを指定する。ここでは、このシグネチャを構成するメンバーが1つだけであるため、その重要度が1.0となっている。

【0255】

そして、特徴データは、同図における29、30、31行目で、それぞれ、シグネチャ・ディスクリプタの終わり及びアブストラクト・ディスクリプタの終わり並びにショット・ディスクリプタの終わりを示す。

【0256】

次に、特徴データは、同図における32行目でショット・ディスクリプタの始

まりを示し、第2番目のこのシーンの構成要素がshot2であることを指定し、その内容を以下に記述することを宣言する。

【0257】

また、特徴データは、同図における33、34行目で、それぞれ、shot2の開始位置及び終了位置を、時間やフレーム番号等により指定する。ここでは、開始位置がフレーム番号“12617”であり、終了位置がフレーム番号“12628”であることを示している。ディスクリプションスキーム上でのこの情報は、図41中のコンポーネントを表すノードの一段だけ下位のノードにおいてスタートタイム(startTime)及びエンドタイム(endTime)により表される。

【0258】

また、特徴データは、図42における35行目の位置に、shot2の内容を示す多くのディスクリプタが記述されるが、ここでは、省略する。

【0259】

そして、特徴データは、同図における36行目でショット・ディスクリプタの終わりを示す。

【0260】

また、特徴データは、同図における37、38行目の位置に、それぞれ、多くのshotの内容を示す多くのディスクリプタと、各shotを記述する多くのディスクリプタが記述されるが、ここでは、省略する。

【0261】

そして、特徴データは、同図における39行目でコンポーネント・ディスクリプタの終わりを示す。

【0262】

次に、特徴データは、同図における40行目でこのシーンのインデックス(index)を示す幾つかのAVハイアラキ(AVHierarchy)・ディスクリプタが以下に記述されることを宣言する。

【0263】

また、特徴データは、同図における41行目でハイアラキ(hierarchy)・ディスクリプタの始まりを示し、このハイアラキがショット・インデックス(

SHOT_INDEX) で示されることを宣言する。

【0264】

さらに、特徴データは、同図における42行目でクラスタ (cluster) ・ディスクリプタの始まりを示し、第1番目のクラスタを以下に記述し、このクラスタがcluster1という名前であることを宣言する。

【0265】

次に、特徴データは、同図における43行目でcluster1のメンバーの指定開始を示す。

【0266】

また、特徴データは、同図における44行目でcluster1のメンバーがショットであり、名前がshot2、shot4、・・・であることを指定する。

【0267】

さらに、特徴データは、同図における45行目でcluster1のメンバーの指定終了を示す。

【0268】

次に、特徴データは、同図における46行目でアブストラクト・ディスクリプタの始まりを示し、cluster1の内容を代表する部分の指定を行う。

【0269】

また、特徴データは、同図における47行目でプロトタイプ (prototype) ・ディスクリプタを示す。ここでは、cluster1の内容を代表する部分がshot4であることを示している。

【0270】

さらに、特徴データは、同図における48、49行目で、それぞれ、アブストラクト・ディスクリプタの終わり及びクラスタ・ディスクリプタの終わりを示す。

【0271】

次に、特徴データは、同図における50、51行目で、それぞれ、クラスタ・ディスクリプタを示し、第2、3番目のクラスタがcluster2、clus

ter3という名前であることを宣言する。なお、これらの行の位置には、その内容を記述する多くのディスクリプタが示されるが、ここでは、省略する。

【0272】

そして、特徴データは、同図における52行目でハイアラキ・ディスクリプタの終わりを示す。

【0273】

また、特徴データは、同図における53行目でこのシーンのインデックスを示す幾つかのAVハイアラキの記述が終わることを宣言する。

【0274】

さらに、特徴データは、同図における54行目の位置に、このシーンの内容を示す多くのディスクリプタが記述されるが、ここでは、省略する。

【0275】

次に、特徴データは、同図における55行目でこのシーンの構造を示す幾つかのAVグラフ (AVGraph) ・ディスクリプタが以下にあることを宣言する。

【0276】

そして、特徴データは、同図における56行目でAVグラフ・ディスクリプタの始まりを示し、このグラフがショット・グループ (SHOT_GROUP) であることを宣言する。

【0277】

また、特徴データは、同図における57乃至61行目でAVグラフ縁 (AVGraphEdge) ・ディスクリプタを示す。57行目では、グラフの開始点がcluster1であることを示している。このcluster1は、42行目で指定されている。

【0278】

さらに、特徴データは、同図における58、59、60行目で、それぞれ、cluster1からcluster2に、cluster3からcluster2に、cluster2からcluster3にグラフが伸びていることを示している。

【0279】

そして、特徴データは、同図における61行目でグラフの終了点がcluster 3であることを示し、62行目でAVグラフ・ディスクリプタの終わりを示す。

【0280】

また、特徴データは、同図における63行目でこのシーンの構造を示す幾つかのAVグラフ・ディスクリプタの記述をこれで終わることを宣言する。

【0281】

そして、特徴データは、同図における64行目の位置で、このシーンの内容を記述する多くのディスクリプタを示した後、65行目でシーン記述を示すディスクリプタの終わりを示す。

【0282】

このようにして、あるビデオデータにおけるシーンストラクチャを記述するディスクリプションスキームは、複数のディスクリプタの集合により表され、特徴データを記述することができる。

【0283】

送信装置10は、特徴データ生成部12によりビデオデータに応じたこのような特徴データを生成し、ビデオデータとともに受信装置20へと送信する。そのため、ビデオデータとともに特徴データを受信した受信装置20は、特徴データに基づき、ビデオデータを解析するだけで、ビデオデータのビデオストラクチャを容易に回復することができ、検索エンジン23により例えばストーリーボードを作成することができる。

【0284】

したがって、受信装置20は、ビデオデータのビデオストラクチャを解析するための高度な処理能力や大容量の計算資源を有する必要がなく、低コスト化及び小容量メモリ化を図ることができる。

【0285】

そして、ユーザは、事前にビデオデータの内容のストラクチャを知る必要はなく、ビデオデータに対して容易でかつ個人的にアクセスすることが可能となり、

ブラウジングを効率よく行うことができる。

【0286】

以上のように、ディスクリプションスキームとして定義される本発明にかかるデータ記述方法は、ビデオデータの内容を検索してブラウジングする高度な方法を実現するものであって、このデータ記述方法により記述された特徴データは、ビデオデータのビデオストラクチャを的確にかつ簡便に効率よく表すことができる。また、上位クラスのディスクリプタの機能を継承してディスクリプタを構成することから、ディスクリプタが高度な機能を有するものへと容易に拡張され得て拡張性に優れたディスクリプションスキームを構築することが可能であるとともに、特徴データの容量を小さくすることができ、受信装置20の小容量メモリ化に対する一助となる。

【0287】

なお、本発明は、上述した実施の形態に限定されるものではなく、例えば、ディスクリプタは、上位のディスクリプタの機能が継承されていなくてもよい。ただしこの場合、特徴データには、1つのディスクリプタに関して、全ての上位のディスクリプタが書き下される必要がある。

【0288】

また、本発明は、ビデオデータからビデオストラクチャを抽出する処理としては、上述したもの以外であってもよいことは勿論である。

【0289】

このように、本発明は、その趣旨を逸脱しない範囲で適宜変更が可能であることはいふまでもない。

【0290】

【発明の効果】

以上詳細に説明したように、本発明にかかるデータ記述方法は、入力データの内容の特徴を表す特徴データを記述するためのデータ記述方法であって、記述方法を定義する単位要素となる記述子は、下位要素としての属性が定義される構造を有し、属性として記述子を含み得る。

【0291】

したがって、本発明にかかるデータ記述方法は、構造を有する記述子の集合を用いて構造化された特徴データを記述することができ、この特徴データに基づいて、容易に入力データの内容の特徴を抽出することができる。

【0292】

また、本発明にかかるデータ処理装置は、入力データの内容の特徴を表す特徴データを生成するデータ処理装置であって、記述方法を定義する単位要素となる記述子であり、下位要素としての属性が定義される構造を有し、属性として記述子を含み得る記述子から特徴データを生成する生成手段を備える。

【0293】

したがって、本発明にかかるデータ処理装置は、構造を有する記述子の集合を用いて構造化された特徴データを生成することができ、この特徴データに基づいて、的確かつ簡便に効率よく入力データの内容の特徴を表現することができる。

【0294】

さらに、本発明にかかるデータ処理装置は、入力データの内容の特徴を表す特徴データを利用するデータ処理装置であって、記述方法を定義する単位要素となる記述子であり、下位要素としての属性が定義される構造を有し、属性として記述子を含み得る記述子から生成された特徴データを利用して、入力データの内容の特徴を復元する復元手段を備える。

【0295】

したがって、本発明にかかるデータ処理装置は、特徴データに基づいて、入力データの内容の特徴を復元することができ、高度な処理能力や大容量の計算資源を必要とせずに入力データの内容の特徴を抽出することができる。

【図面の簡単な説明】

【図1】

本発明において適用するビデオデータの構成を説明する図であって、モデル化したビデオデータのストラクチャを説明する図である。

【図2】

本発明の実施の形態として示すデータ送受信処理システムの構成を説明するブ

ロック図である。

【図3】

ディスクリプタの記述フォーマットを説明する図である。

【図4】

ディスクリプタの継承関係を説明する図である。

【図5】

AVエレメント・ディスクリプタの記述フォーマットを説明する図である。

【図6】

AVセグメント・ディスクリプタの記述フォーマットを説明する図である。

【図7】

AVアブストラクト・ディスクリプタの記述フォーマットを説明する図である

【図8】

AVプロトタイプ・ディスクリプタの記述フォーマットを説明する図である。

【図9】

プロトタイプを説明する図である。

【図10】

AVハイアラキ・ディスクリプタの記述フォーマットを説明する図である。

【図11】

クラスタ・ディスクリプタの記述フォーマットを説明する図である。

【図12】

クラスタリングを説明する図である。

【図13】

コンポジットを説明する図である。

【図14】

コンポジット・セグメント・ディスクリプタの記述フォーマットを説明する図である。

【図15】

メディア構造グラフを説明する図である。

【図 16】

AV グラフ・ディスクリプタの記述フォーマットを説明する図である。

【図 17】

AV グラフ縁・ディスクリプタの記述フォーマットを説明する図である。

【図 18】

シグネチャ・ディスクリプタの記述フォーマットを説明する図である。

【図 19】

テンポラル・シグネチャ・ディスクリプタの記述フォーマットを説明する図である。

【図 20】

アクティビティ・レベル・ディスクリプタの記述フォーマットを説明する図である。

【図 21】

ウエイト・ディスクリプタの記述フォーマットを説明する図である。

【図 22】

ウエイト・ベクトル・ディスクリプタの記述フォーマットを説明する図である。

【図 23】

カラー・ディスクリプタの記述フォーマットを説明する図である。

【図 24】

カラー・スペース・ディスクリプタの記述フォーマットを説明する図である。

【図 25】

コンセプト・ディスクリプタの記述フォーマットを説明する図である。

【図 26】

カラー・シグネチャ・ディスクリプタの記述フォーマットを説明する図である。

【図 27】

シノプシス・ディスクリプタの記述フォーマットを説明する図である。

【図28】

音声フレーム・ディスクリプタの記述フォーマットを説明する図である。

【図29】

映像フレーム・ディスクリプタの記述フォーマットを説明する図である。

【図30】

ブラウジング・ノード・ディスクリプタの記述フォーマットを説明する図であ

る。

【図31】

ショット・ディスクリプタの記述フォーマットを説明する図である。

【図32】

シーン・ディスクリプタの記述フォーマットを説明する図である。

【図33】

トピック・ディスクリプタの記述フォーマットを説明する図である。

【図34】

プログラム・ディスクリプタの記述フォーマットを説明する図である。

【図35】

ハイライト・ディスクリプタの記述フォーマットを説明する図である。

【図36】

上記データ送受信処理システムにおける送信装置において、シーンを検出してグループ化する際の一連の工程を説明するフローチャートである。

【図37】

上記データ送受信処理システムにおける送信装置における動的フィーチャサンプリング処理を説明する図である。

【図38】

非類似性閾値を説明する図である。

【図39】

時間閾値を説明する図である。

【図40】

上記データ送受信処理システムにおける送信装置において、セグメントをグル

ー普化する際の一連の工程を説明するフローチャートである。

【図 4 1】

ディスクリプションスキームの一例を説明する図である。

【図 4 2】

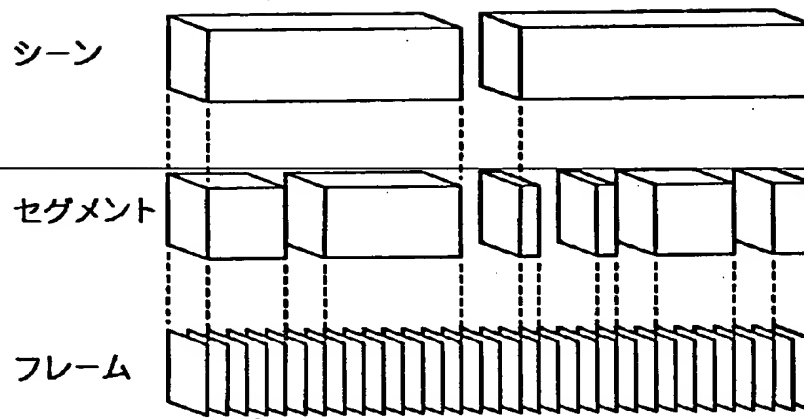
特徴データの一例を説明する図である。

【符号の説明】

10 送信装置、 11 記憶部、 12 特徴データ生成部、 20 受信
装置、 23 検索エンジン

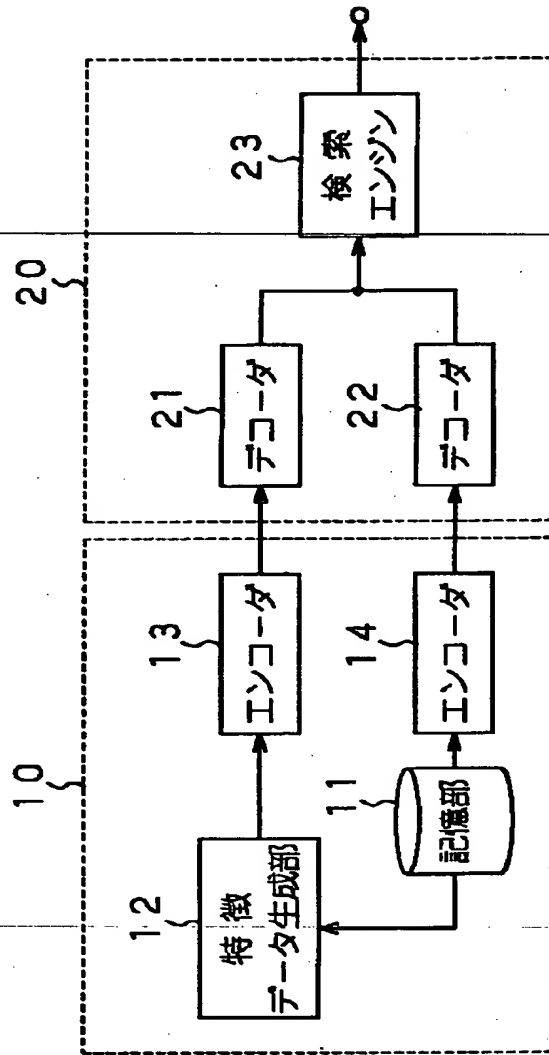
【書類名】 図面

【図 1】



ビデオストラクチャの階層モデル

【図2】



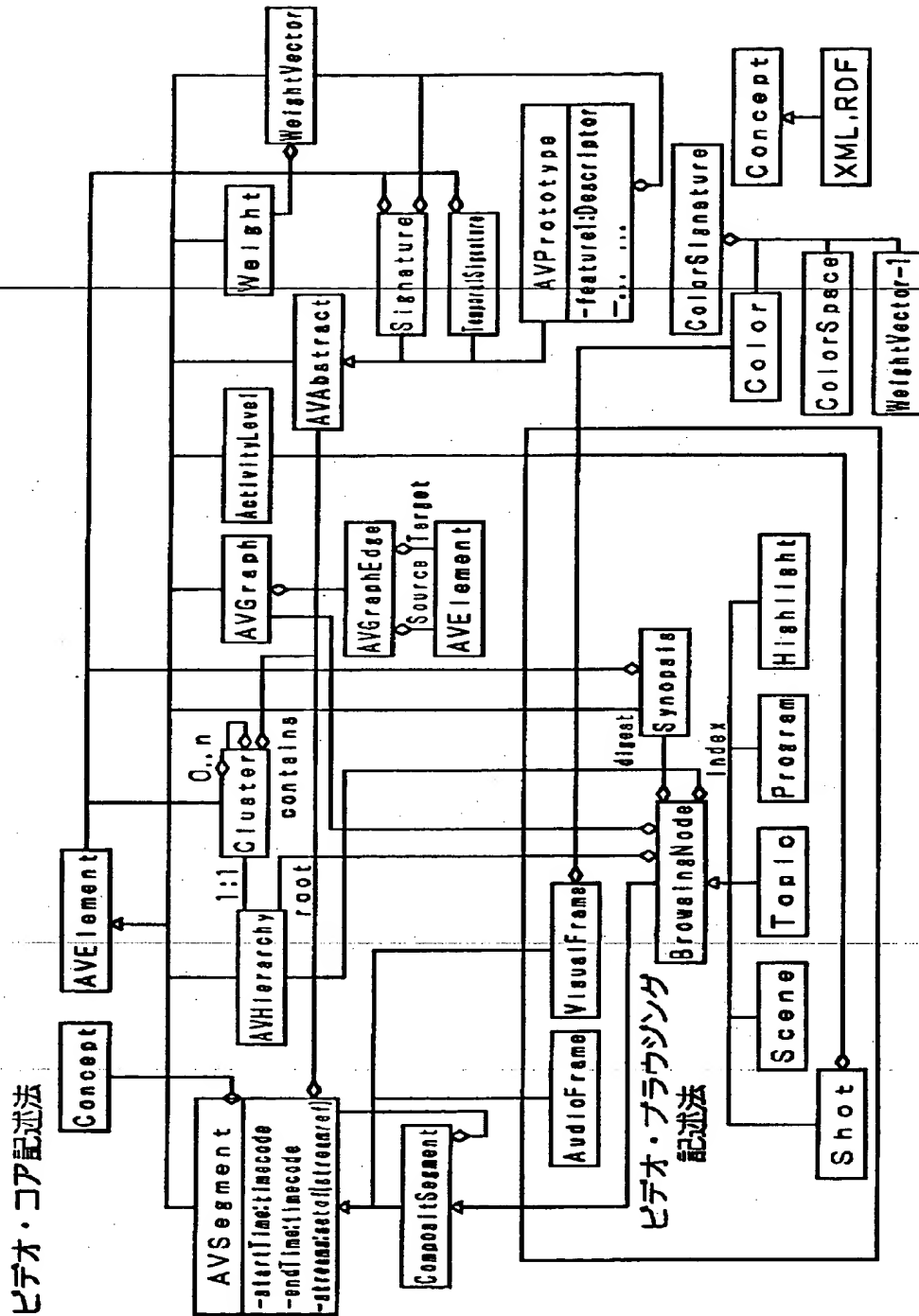
データ送受信処理システムの構成ブロック図

【図3】

Descriptor		Superclass
Feature	Type	Example Value
⋮	⋮	⋮
⋮	⋮	⋮

ディスクリプタの記述フォーマット

【図4】



ディスクリプタの継承関係の説明図

【図 5】

Descriptor: AVElement(AVE)		Superclass: なし
Feature:	Feature: Type	Example Value

ディスクリプタの記述フォーマット

【図 6】

Descriptor: AVSegment(AVS)		Superclass: AVElement
Feature:	Feature Type	Example Value:
StartTime	Timecode	
EndTime	Timecode	
Streams	setof(streamref)	
Abstract	setof(AVAbstract)	Key frame
Concept	setof(Concept)	An RDF structure describing the content of this segment

ディスクリプタの記述フォーマット

【図 7】

Descriptor: AVAbstract(ABST)		Superclass: AVElement
Feature:	Feature Type	Example Value:

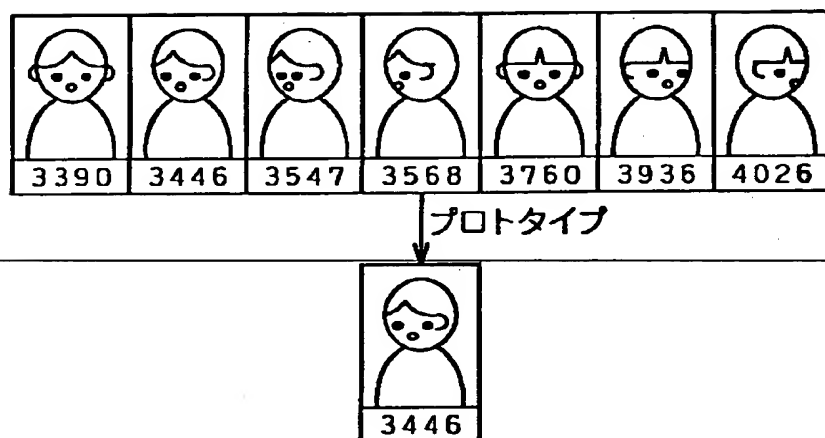
ディスクリプタの記述フォーマット

【図 8】

Descriptor: AVPrototype(AVP)		Superclass: AVAbstract
Feature:	Feature Type	Example Value:
Kind	String	"shot" "scene" "visualframe"
weights	WeightVector	
feature1	Descriptor1	
feature2	Descriptor2	
...
Feature n	Descriptor n	

ディスクリプタの記述フォーマット

【図9】



プロトタイプの説明図

【図 10】

Descriptor: AVHierarchy(HIER)		Superclass: AVElement
Feature:	Feature Type	Example Value:
Root	Cluster	
Kind	String	

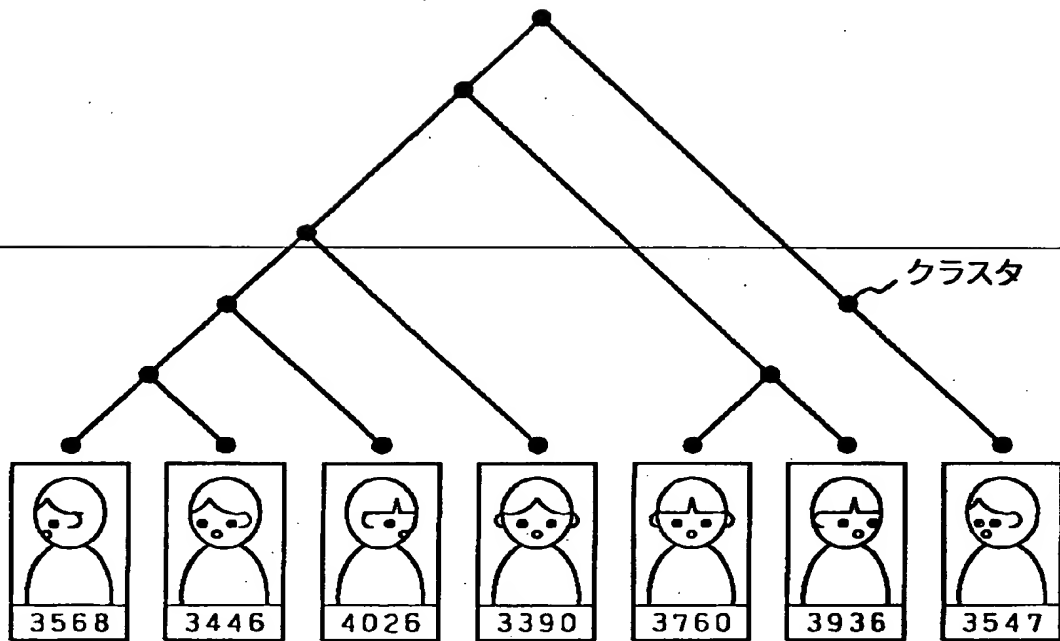
ディスクリプタの記述フォーマット

【図 11】

Descriptor: Cluster(CN)		Superclass: AVElement
Feature:	Feature Type	Example Value:
members	setof(AVElement)	
abstract	AVAAbstract	

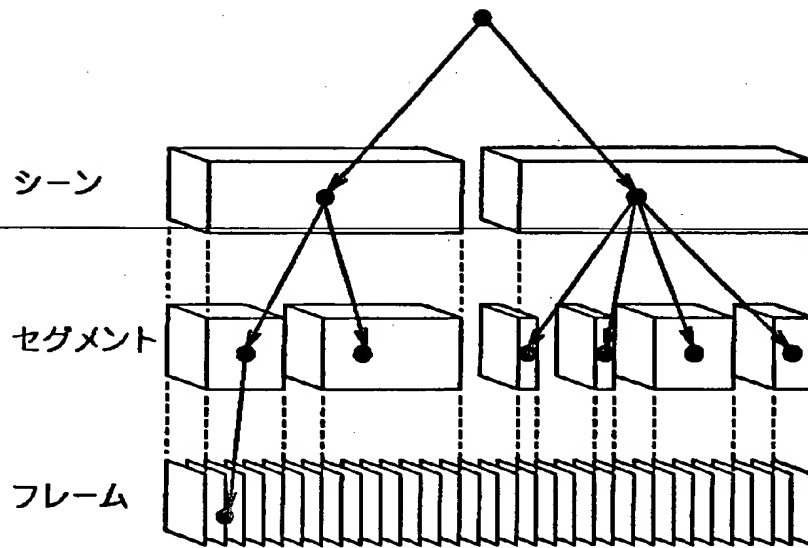
ディスクリプタの記述フォーマット

【図 12】



クラスタリングの説明図

【図 13】



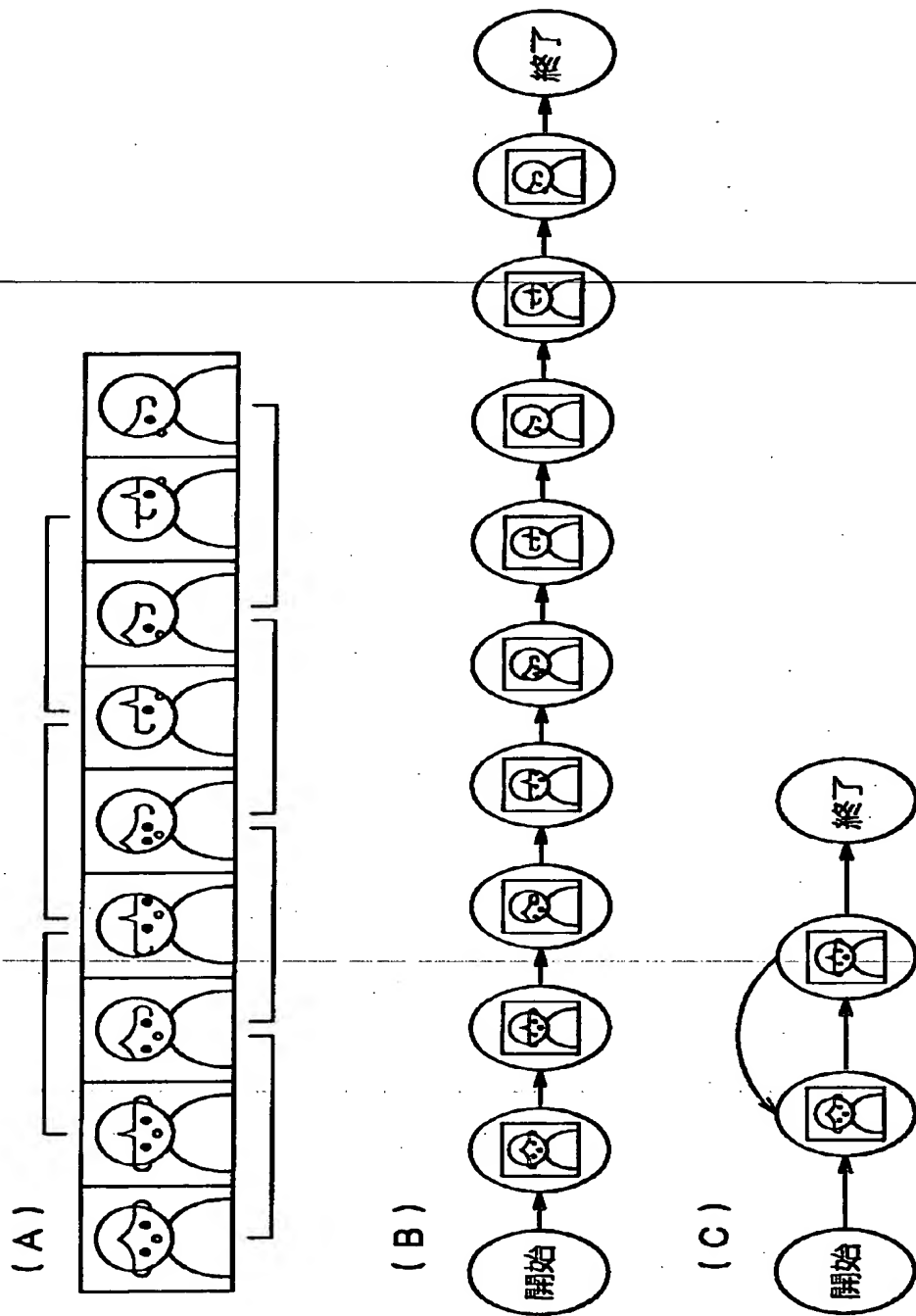
コンポジットの説明図

【図 14】

Descriptor: CompositeSegment(SN)		Superclass: AVSegment
Feature:	Feature Type	Example Value:
components	seqof(AVSegment)	

ディスクリプタの記述フォーマット

【図 15】



メディア構造グラフの説明図

【図 16】

Descriptor: AVGraph(AVG)		Superclass: AVElement
Feature:	Feature Type	Example Value:
Edges	setof(AVGraphEdge)	
Kind	string	

ディスクリプタの記述フォーマット

【図 17】

Descriptor: AVGraphEdge(AVEDGE)		Superclass: なし
Feature:	Feature Type	Example Value:
Source	AVElement	
Target	AVElement	
Label	string	"follows" "similarTo"

ディスクリプタの記述フォーマット

【図 18】

Descriptor: Signature(SIG)		Superclass: AVAbstract
Feature:	Feature Type	Example Value:
Weights	WeightVector[n]	
members	AVElement[n]	
Kind	string	

ディスクリプタの記述フォーマット

【図 19】

Descriptor: TemporalSignature(TS)		Superclass: AVAbstract
Feature:	Feature Type	Example Value:
time	WeightVector	
members	Seqof(AVElement)	
Kind	string	

ディスクリプタの記述フォーマット

【図 20】

Descriptor: ActivityLevel(ACT)		Appliesto AVElement	
Feature:	Feature Type	Example Value:	
ActivityLevel	Real		

ディスクリプタの記述フォーマット

【図 21】

Descriptor: Weight(WGT)		Appliesto AVElement	
Feature:	Feature Type		
Weighting	Real		

ディスクリプタの記述フォーマット

【図 2 2】

Descriptor: WeightVector(WV)		Appliesto AVElement
Feature:	Feature Type	Example Value:
Weights	Weight[n]	
normalized	Boolean	

ディスクリプタの記述フォーマット

【図 2 3】

Descriptor: Color(COL)	Appliesto: AVElement	Super type: なし
Feature:	Feature Type	Example Value:
color	Real[3]	

ディスクリプタの記述フォーマット

【図 24】

Descriptor: ColorSpace(CS)	Applies to: AVElement	Super type: なし
Feature:	Feature Type	Example Value:
Colorspace	String	"HSV"

ディスクリプタの記述フォーマット

【図 25】

Descriptor: Concept(CPT)	Applies to: AVElement	Super type: なし
Feature:	Feature Type	Example Value:
concept	RDFNode or XML or ...	

ディスクリプタの記述フォーマット

【図 26】

Descriptor: ColorSignature(CHT)	Applies to: AVElement	Super type: なし
Feature:	Feature Type	Example Value:
Colorspace	Colorspace	"RGB"
Colors	Color[n]	
Weights	WeightVector	

ディスクリプタの記述フォーマット

【図 27】

Descriptor: Synopsis	Superclass: AVElement	
Feature: Concept(CPT)	Feature Type	Example Value:
Kind	String	SLIDE SHOW
Sequence	Listof(AVElement)	
IsOrdered	Boolean	

ディスクリプタの記述フォーマット

【図 28】

Descriptor: AudioFrame(AF)	Superclass: AVSegment
Feature:	Feature Type

ディスクリプタの記述フォーマット

【図 29】

Descriptor: VisualFrame(VF)		Superclass: AVSegment
Feature:	Feature Type	Example Value:
Color	ColorSignature	

ディスクリプタの記述フォーマット

【図 30】

Descriptor: BrowsingNode(BRO)		Superclass: CompositeSegment
Feature:	Feature Type	Example Value:
Index	Setof(AVHierarchy)	A classification of the shots in a video
Synopsis	Setof(AVSynopsis)	Sequence of key frames
Structure	AVGraph(optional)	

ディスクリプタの記述フォーマット

【図 31】

Descriptor: Shot(SHT)		Superclass: BrowsingNode
Feature:	Feature Type	Example Value:
Activity	ActivityLevel	"0.5"
Abstract	From AVSegment	
Synopsis	From BrowsingNode	
Index	From BrowsingNode	
Structure	From BrowsingNode	

ディスクリプタの記述フォーマット

【図 3 2】

Descriptor: Scene(SCN)		Superclass: BrowsingNode
Feature:	Feature Type	Example Value:
Kind	String	"DIALOG" "COMMERICAL" "ACTION" "ANCHOR"
Abstract	From AVSegment	
Synopsis	From BrowsingNode	
Index	From BrowsingNode	
Structure	From BrowsingNode	

ディスクリプタの記述フォーマット

【図 33】

Descriptor: Topic(TPC)		Superclass: BrowsingNode
Feature:	Feature Type	Example Value:
Abstract	From AVSegment	
Synopsis	From BrowsingNode	
Index	From BrowsingNode	
Structure	From BrowsingNode	

ディスクリプタの記述フォーマット

【図 34】

Descriptor: Program(PRG)		Superclass: BrowsingNode
Feature:	Feature Type	Example Value:
Abstract	From AVSegment	Kind: TOPIC ABSTRACT SCENE ABSTRACT SHOT ABSTRACT
Synopsis	From BrowsingNode	
Index	From BrowsingNode	
Structure	From BrowsingNode	

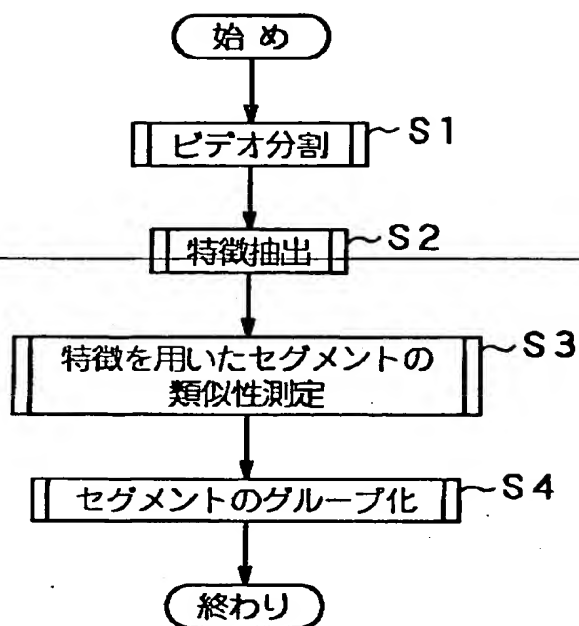
ディスクリプタの記述フォーマット

【図 35】

Descriptor: Highlight(HI)		Superclass: BrowsingNode
Feature:	Feature Type	Example Value:
Interest	Weight	

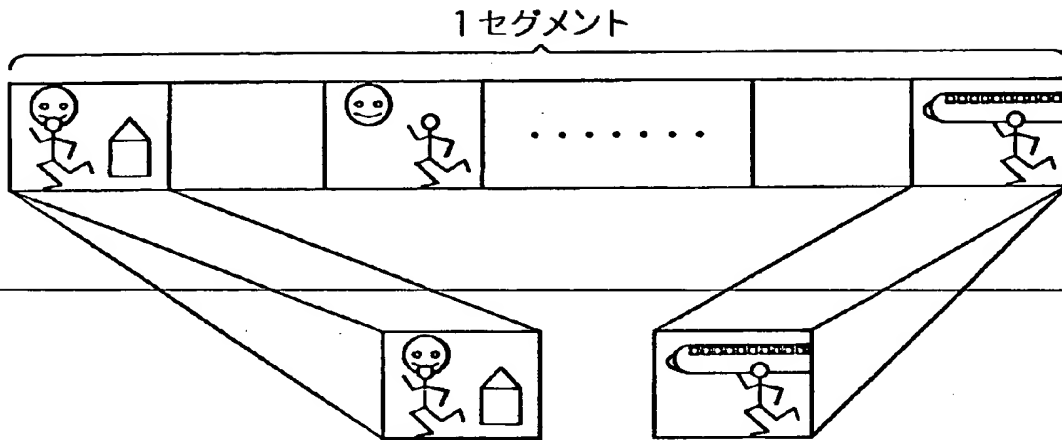
ディスクリプタの記述フォーマット

【図 36】



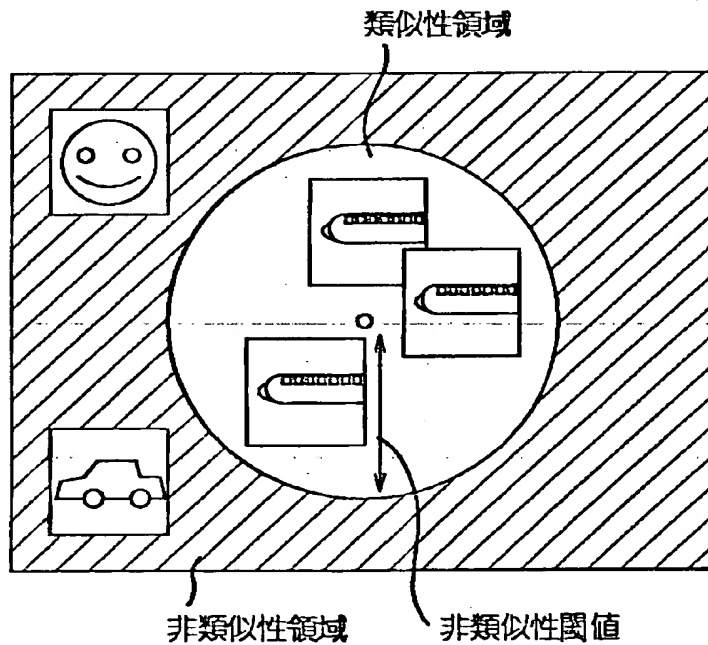
送信装置における一連の処理工程

【図 37】



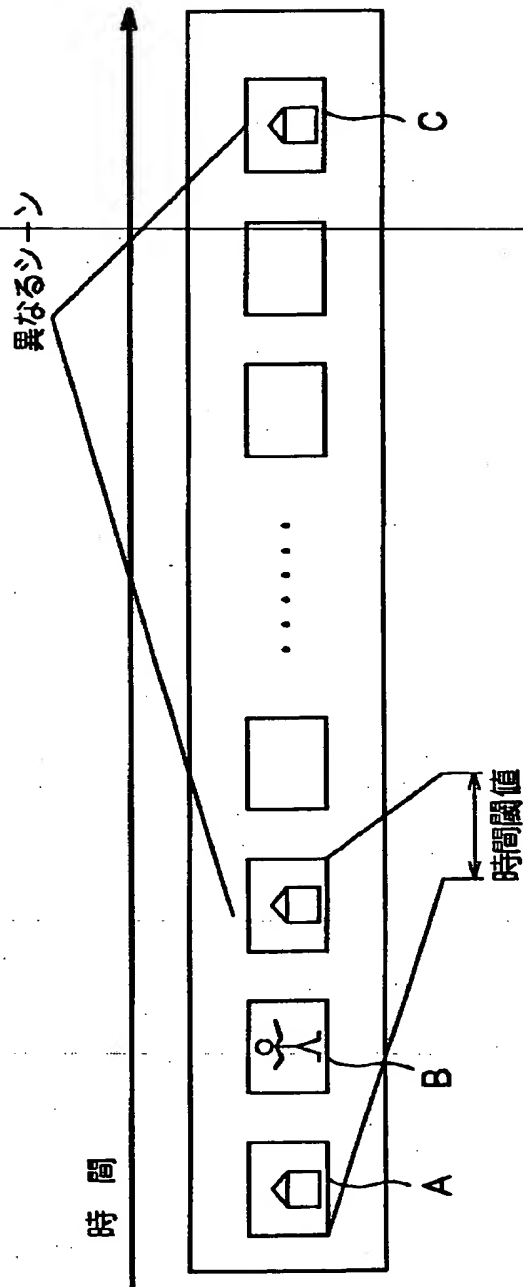
動的特徴のサンプリング方法の説明図

【図 38】



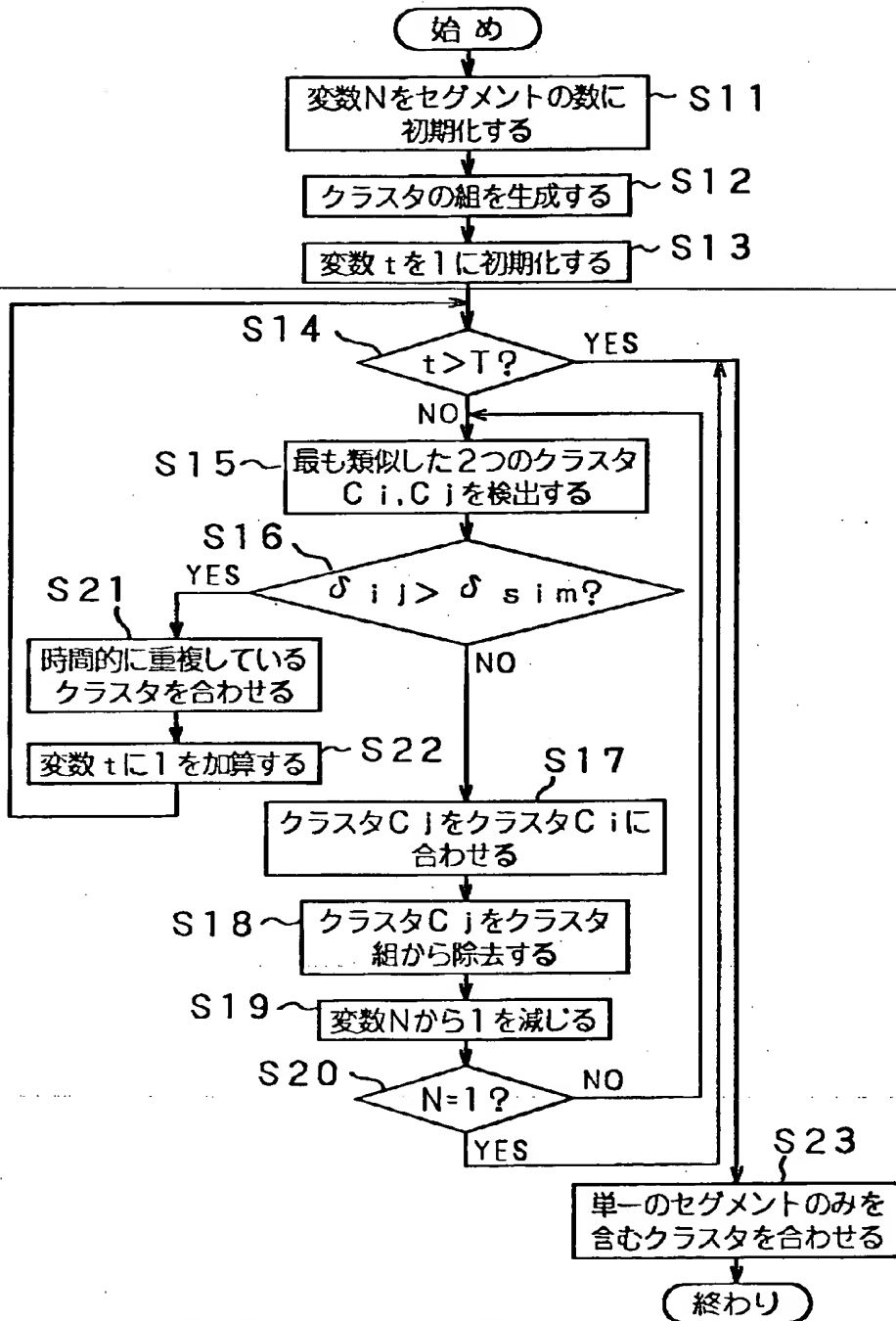
非類似性閾値の説明図

【図 39】



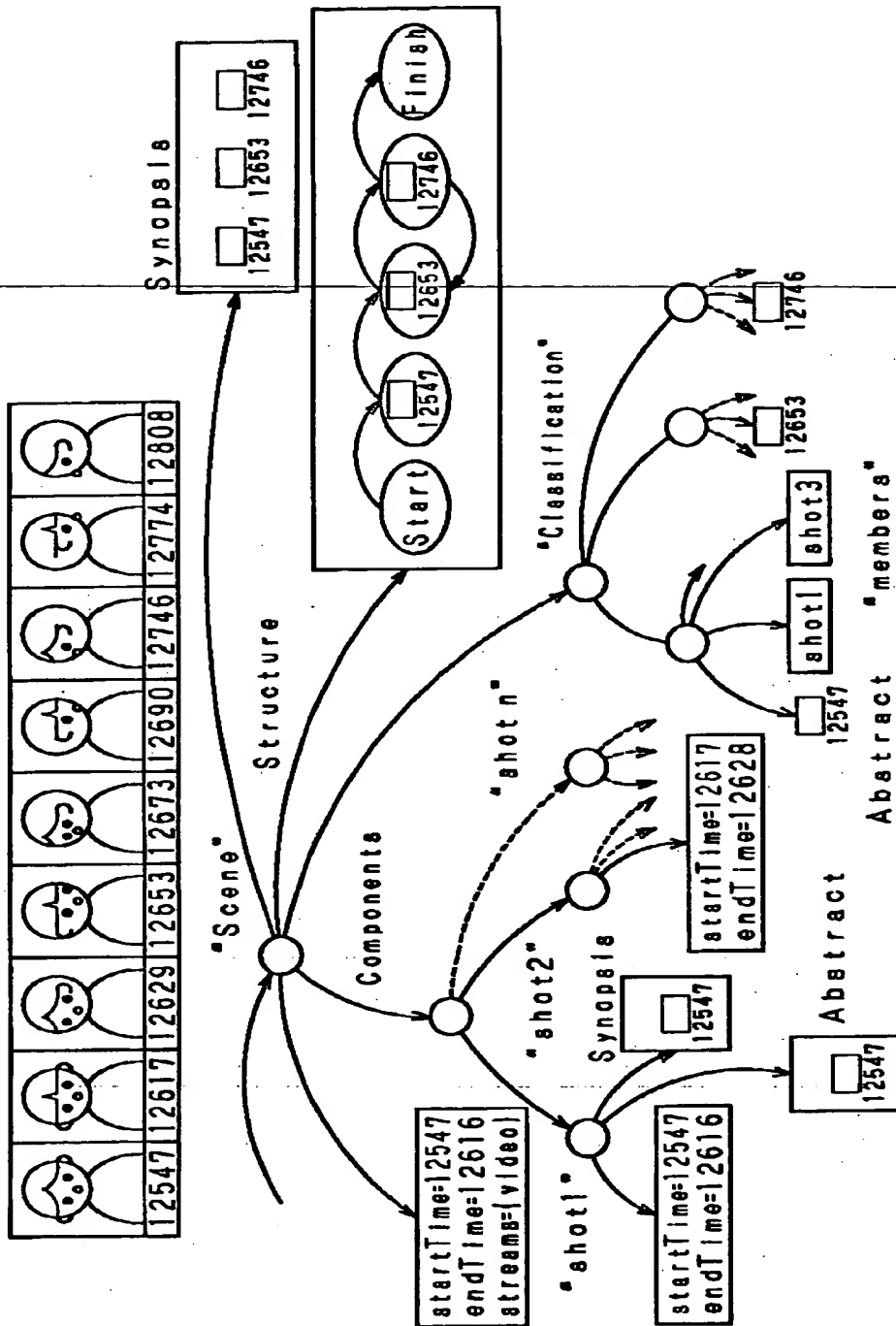
時間閾値の説明図

【図 40】



送信装置における一連の処理工程

【図41】



ディスクリプションスキームの一例

【図 4 2】

```

<scene>
  <startPoint>12547</startPoint> ; AVSegment
  <endPoint>12830</endPoint> ; AVSegment
  <streams>{video}</streams> ; AVSegment
  <synopsis kind=KEY_FRAME> ; BrowsingNode
    <sequence>
      < KEY_FRAME refid= KEY_FRAME 1>
      < KEY_FRAME refid= KEY_FRAME 2>
      < KEY_FRAME refid= KEY_FRAME 3>
    </sequence>
  </synopsis>
  <abstract> ; AVSegment
    <signature kind=SHOT_SIGNATURE>
      <members>
        <shot refid=shot1> <shot refid=shot2>
        <shot refid=shot3>
      </members>
      <weights>0.1 0.45 0.45</weights>
    </signature>
  </abstract>
  <components> ; CompositeSegment
    <shot id=shot1> ; Shot
      <startPoint>12547</startPoint> ; AVSegment
      <endPoint>12616</endPoint> ; AVSegment
      <abstract> ; AVSegment
        <signature kind=FRAME_SIGNATURE>
          <members><frame refid=12590></members>
          <weights>1.0</weights>
        </signature>
      </abstract>
    </shot>
    <shot id=shot2>
      <startPoint>12617</startPoint>
      <endPoint>12628</endPoint>
    </shot>
    =
    <shot id=shotn></shot>
  </components>
  <index> ; BrowsingNode ; Hierarchy
    <hierarchy kind=SHOT_INDEX>
      <cluster id=cluster1>
        <members>
          <shot refid=shot2> <shot refid=shot4>...
        </members> ; Cluster
        <abstract>
          <prototype> shot 4</prototype>
        </abstract>
      </cluster>
      <cluster id=cluster2></cluster>
      <cluster id=cluster3></cluster>
    </hierarchy>
  </index>
  =
  <structure> ; BrowsingNode ; AVGraph
    <avgraph kind=SHOT_GROUP>
      <avgraphedge source=start target=cluster1>
      <avgraphedge source=cluster1 target=cluster2>
      <avgraphedge source=cluster3 target=cluster2>
      <avgraphedge source=cluster2 target=cluster3>
      <avgraphedge source=cluster3 target=finish>
    </avgraph>
  </structure>
  =
</scene>

```

特徴データの一例

【書類名】 要約書

【要約】

【課題】 種々のビデオデータの内容を検索してブラウジングする高度な方法を実現する。

【解決手段】 ビデオデータの内容の特徴を表す特徴データを記述するための記述方法を定義する単位要素となるディスクリプタは、下位要素としての属性が定義される構造を有し、属性として他のディスクリプタを含み得る記述フォーマットにより形成される。すなわち、ディスクリプタの記述フォーマットは、下位要素のディスクリプタ或いは数値の種類を示すフィーチャ (Feature) と、下位要素のディスクリプタ或いは数値のタイプを示すフィーチャタイプ (Feature Type) と、下位要素のディスクリプタ或いは数値の例を示す実例値 (Example Value) とからなる。

【選択図】 図3

認定・付加情報

特許出願の番号	平成11年 特許願 第023068号
受付番号	59900079347
書類名	特許願
担当官	第一担当上席 0090
作成日	平成11年 2月13日

<認定情報・付加情報>

【提出日】 平成11年 1月29日

出 願 人 履 歴 情 報

識別番号

[000002185]

1. 変更年月日 1990年 8月30日

[変更理由] 新規登録

住 所 東京都品川区北品川6丁目7番35号

氏 名 ソニー株式会社

This Page Blank (uspto)